Structural and Biochemical Studies of Cohesin Regulation by Wpl1

Avradip Chatterjee

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Dr. Martin Singleton

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2012

Declaration

I Avradip Chatterjee confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Sister chromatid cohesion physically tethers the sister chromatids formed after DNA replication ensuring biorientation and proper segregation of chromosomes. Cohesion is mediated by a ring shaped protein complex called cohesin, formed of the conserved subunits – Smc1, Smc3, Scc1 and Scc3 (in yeast). Additionally, several regulatory proteins control the cohesin complex, important among which are Eco1 and Wpl1. These two proteins have been shown to have opposing functions, wherein Wpl1 tends to maintain cohesin in a noncohesive state whereas Eco1 counteracts this effect by acetylating cohesin, which switches the complex to a cohesive state. Nevertheless, the detailed mechanism of functioning of these proteins and as a result, the process of cohesion establishment and regulation as a whole still remains poorly understood.

The present work describes characterization of the protein Wpl1 using structure-function studies in order to understand the process of cohesin regulation in further detail. The conserved domain of Wpl1, called WAPL, was found to directly interact with the ATPase head of the Smc3 subunit of the cohesin complex. The interaction has been verified using various biophysical and biochemical methods. In addition, crystal structures of the WAPL domain alone and in complex with an interacting stretch of the Smc3 have been described. Together with the complementary functional studies, the structures reveal novel features and probable mechanism by which Wpl1 may regulate sister chromatid cohesion. Thus, the work has provided mechanistic insights into the functioning of the protein and an improvement in the understanding of regulation of cohesion.

Acknowledgements

First and foremost I would like to express my sincere gratitude to Dr. Martin Singleton who has been a great mentor and friend and has guided, advised, helped and encouraged me through all the ups and downs for the last four years. I would also like to take this opportunity to thank the members of my thesis committee, Dr. Frank Uhlmann and Dr. Peter Parker, for valuable suggestions and guidance whenever needed.

I deeply acknowledge the selfless efforts of Dr. Alan Purvis who cared to take out time from his busy schedule to demonstrate and teach various aspects of crystallograhy and am grateful to Dr. Xiao Wen Hu and Dr. Silva Zakian for extending their help at various stages of my project. I would also like to thank all my other lab members, past and present, for their constant co-operation and support.

I am grateful to the London Research Institute and Cancer Research UK for giving me the opportunity and platform to learn and work in my chosen field of interest and also contribute toward a noble cause.

Last but not the least, I am thankful to God and each of my family members and friends who have been my constant source of support, inspiration and encouragement.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP-binding cassette
ATP	Adenosine triphosphate
BLI	Biolayer interferometry
βME	beta-Mercaptoethanol
DNA	Deoxyribinucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
FP	Fluorescence polarisation
GST	Glutathione S-transferase
HEPES	Hydroxyethyl piperazineethanesulfonic acid
LB	Luria Bertani medium
MR	Molecular replacement
NBD	Nucleotide binding domain
NHEJ	Non-homologous end joining
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecyl sulphate
TEV	Tobacco etch virus protease
TCEP	Tris(2-carboxyethyl)phosphine

Chapter 1. Introduction

The cell cycle entails a chain of precisely orchestrated events with an ultimate objective of efficient transmission of genetic information to the daughter cells. Cells first need to generate an exact replica of the entire genome, achieved through DNA replication during the S phase. Replication gives rise to pairs of duplicated chromosomes that eventually segregate into the daughters as the cell goes through the events of mitosis or meiosis during M phase. However, the S phase is followed by G2, needed for the cells to get equipped for the subsequent M phase, creating a temporal lapse between the duplication of the chromosomes and their ultimate segregation. This poses an additional challenge for the cells to get over – that of keeping the duplicated chromosomes called the sister chromatids together, through G2 and part of M phase (until anaphase onset), before they can separate from each other. The process of maintaining sister chromatids physically linked, termed sister chromatid cohesion, is one of the most crucial events determining the fidelity of chromosome segregation. Cohesion opposes the poleward forces of the spindle microtubules attached to the kinetochores ensuring amphitelic arrangement of the chromosomes on the metaphase plate, a process termed as biorientation. Following biorientation, the spindle checkpoint is inactivated and cohesion is dissolved allowing progression of the cells to anaphase. Cohesion defects may lead to improper segregation of chromosomes giving rise to genetic disorders and diseases like cancer.

The nature of the links which hold sister chromatids together after DNA replication until anaphase onset has been an area of active research and debate in the past. The observation that DNA replication gives rise to intertwined sister strands where replication forks meet (Sundin and Varshavsky, 1980), initiated the theory that catenation was the principal mechanism to keep sister chromatids linked (Murray and Szostak, 1985). The catenation theory was further supported by genetic studies demonstrating segregation defects in topoisomerase II mutant yeast cells (DiNardo et al., 1984). The idea of catenation giving rise to sister pairing was thoroughly interrogated in a subsequent series of genetic, cell biology and biochemical studies. Results of these showed that in metaphase arrested cells, sister chromatid cohesion is still maintained

although no intertwining could be observed (Guacci et al., 1994; Koshland and Hartwell, 1987). In addition, in topoisomerase II mutants, it was shown that breakage in segregated chromosomes (due to absence of the enzyme) in such cells are restricted only to longer chromosomes arms (with breaks located at ~200kb from centromeres) while the shorter arms were intact following segregation (Spell and Holm, 1994). These observations showed that catenation is not a prerequisite for sister pairing and suggested the existence of a more active mechanism of cohesion. Subsequently, the involvement of a specialized mechanism was supported by the discovery of several proteins essential for sister chromatid cohesion although these were not directly involved in catenation. These studies together demonstrate that although catenation might form passive links between the sister DNA molecules, they are not sufficient to bring about cohesion and identification of several proteins with a more direct role in sister pairing further substantiated the idea.

1.1 Discovery of the cohesin complex

Genes encoding proteins involved in sister chromatid cohesion were identified using genetic screens for mutants in *Drosophila* and yeast. Initially, two genes to be implicated in sister cohesion using these screens were the Drosophila *mei-S332* and *ord*, both of which were required for meiotic sister chromatid cohesion (Davis, 1971) (Kerrebrock et al., 1992) (Miyazaki and Orr-Weaver, 1992). The *mei-S332* mutants showed precocious separation of sister centromeres during anaphase I. The gene product appeared necessary for the sister chromatids to stay together starting from late anaphase I until metaphase II. However, the *ord* mutants showed cohesion defects much earlier, starting from prometaphase I. Characterization of the *mei-S332* and *ord* mutants provided the first evidence for the recruitment of special factors for sister cohesion and also the impetus to identify additional genes. Later, similar genetic screens in budding yeast initially identified *SCC1*, *SCC2*, *SMC1* and *SMC3* while *SCC3*, *SCC4* and *CTF7 (*or *ECO1)* were found later on as the genes important for sister chromatid cohesion during mitosis (Guacci et al., 1997a; Rowland et al., 2009). Importantly, all the genes identified encoded for conserved eukaryotic proteins. The *SCC1* gene product was found to be similar to an earlier identified fission yeast protein called *Rec8*, important

for meiotic cohesion. These studies not only identified dedicated and conserved factors required for sister chromatid cohesion but also showed that a common mechanism existed to physically link the duplicated DNA molecules both during mitosis and meiosis.

The observations in the genetic studies were followed up by biochemical experiments showing that a subset of the proteins, including Smc1, Smc3 and Scc1 identified in the initial screens, are part of a complex, formation of which is crucial for cohesion (Gruber et al., 2003; Losada et al., 1998; Sumara et al., 2000). This complex was named 'cohesin' and could associate with chromosomes from late G1 phase in budding yeast and telophase in vertebrate cells until anaphase onset, that is, exactly when cohesion is maintained (Rowland et al., 2009). The involvement of cohesin in sister chromatid cohesion was further verified by subsequent studies which showed cohesion defects if association of cohesin to chromatin was disrupted either by introducing suitable mutations in the subunits (Arumugam et al., 2003) (Weitzer et al., 2003) or by enzymatic cleavage of either Smc or the Scc1 subunits (separase-mediated cleavage of Scc1 or after artificially introducing TEV cleavage sites in these proteins) (Gruber et al., 2003). These experiments provided additional validation of the cohesion specific roles of these proteins and their distinctive involvement in sister chromatid cohesion was quite apparent at least in budding yeast. Interestingly, in a recent study, it was shown that cohesin helps to maintain intertwining in the metaphase chromosomes and has been proposed to slow down decatenation by Topoisomerase II (Farcas et al., 2011).

The additional proteins detected in these screens that were not part of the cohesin complex were shown to be involved in cohesion more indirectly. These have been found to be important to promote interaction of cohesin and DNA by helping in loading of the complex onto DNA (Scc2 and Scc4) (Rowland et al., 2009) (Toth et al., 1999) (Ciosk et al., 2000), by aiding establishment of cohesion (Ctf7 or Eco1) (Williams et al., 2011) (Toth et al., 1999) or by preventing premature separation of sister chromatids during anaphase (Mei-S332) (Davis, 1971).

In vertebrates, although conserved orthologues of the cohesin subunits exist, their importance in sister chromatid cohesion was not quite clear initially. This was due to the fact that immunodepletion of vertebrate cohesin did not result in a complete loss in cohesion (Losada et al., 1998) (Sumara et al., 2000), unlike in the budding yeast. Also, in vertebrate cells, a major fraction of chromosome bound cohesin is removed during prophase, a process termed as prophase pathway, although sister chromatid cohesion persists until anaphase onset. These led to doubts over the role of cohesin being the major mediator of chromatid cohesion in vertebrate cells and more detailed studies involving vertebrate cohesin were undertaken. It was eventually seen that in the vertebrates, a small proportion, about 5-10%, of cohesin, confined mostly to the centromeres, persists beyond prophase in mammalian cells (Hoque and Ishikawa, 2001; Waizenegger et al., 2000) (Hauf et al., 2001). The unusually high stability of centromeric cohesin in these cells was found to be due to the protein Shugoshin1 (Sgo1), which protects the complex against dissociation from chromatin during prophase (Kitajima et al., 2005; McGuinness et al., 2005). Removal of Sgo1 from cell extracts leads to premature dissociation of centromeric cohesin from chromosomes giving rise to loss of cohesion in prometaphase. The role of cohesin in higher eukaryotes was further verified by studying the effect of deleting the *SCC1* gene from Drosophila, chicken and mammalian cultured cells (Sonoda et al., 2001). These cells showed cohesion loss during S and G2 phases, the effects of which were apparent during the prometaphase. On the other hand, when the prophase pathway is inhibited by depletion of proteins needed to mediate the process (discussed later), an increased association of cohesin with the chromosomes causing delayed sister chromatid resolution is observed. The observations described so far established a unified mechanism of sister chromatid cohesion - that mediated by the protein complex cohesin throughout eukaryotes although some differences exist between budding yeast and higher eukaryotes.

1.2 Architecture of the cohesin core

The core cohesin complex is formed of four subunits – Smc1, Smc3, Scc1 and Scc3, orthologues of which exist throughout eukaryotes. Cohesin architecture was initially studied using electron microscopy (EM), which revealed a conserved ring-shaped morphology of the soluble (i.e., chromatin unbound) form of the complex. The observation was subsequently extended to the chromatin-bound fraction as well based on biochemical studies aimed to characterise the interactions among the subunits (Anderson et al., 2002; Losada et al., 1998; Sumara et al., 2000). The cohesin subunits have been subjected to crystallography based structural analyses more recently, which provided detailed insights into both intra- and inter-subunit interactions of the compex (Haering et al., 2002; Haering et al., 2004) (Kurze et al., 2011).

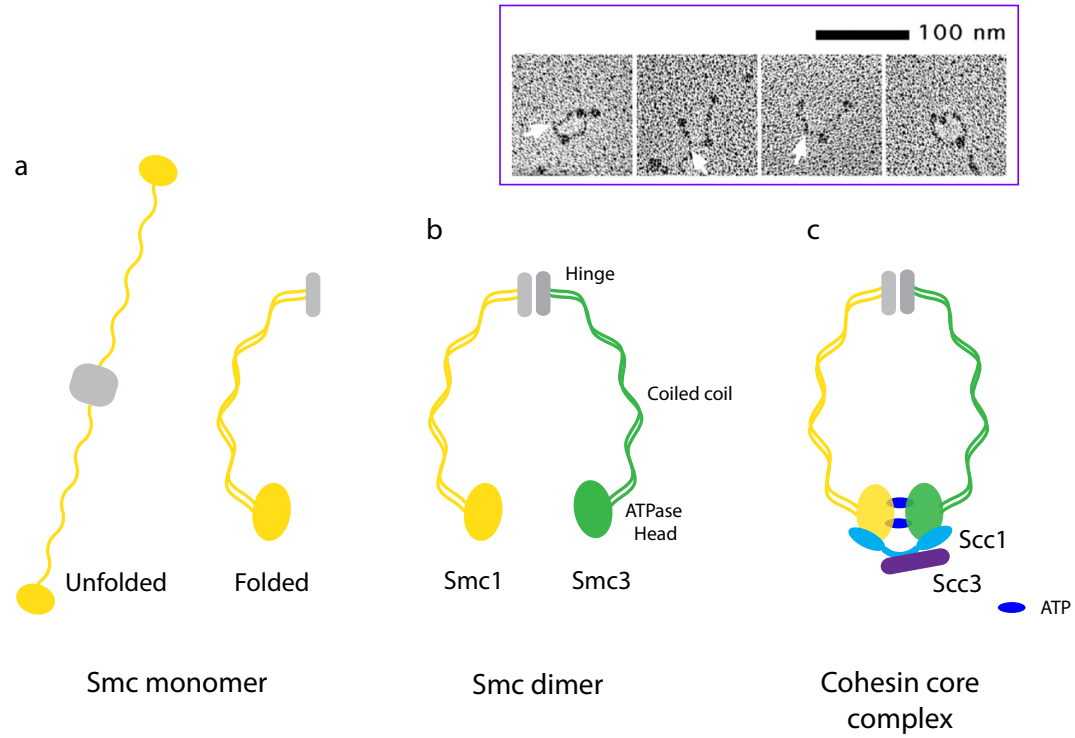


Figure 1-1: Architecture of the cohesin core

Cohesin is a ring shaped complex formed of four conserved subunits Smc1, Smc3, Scc1 and Scc3. (a) Schematic of an Smc monomer in unfolded and folded states, (b) Smc1 and Smc3 proteins forming an Smc heterodimer ; the ATPase head, coiled coil and the hinge domains have been labelled, and (c) the cohesin core complex - the ATPase head dimer sandwich a pair of ATP (dark blue) and the dimerisation is further stabilised by the kleisin subunit Scc1 which also interacts with Scc3, completing the cohesin complex. EM images (Anderson et al., 2002) of human cohesin are shown in the inset. The arrows indicate a kink in one of the coiled coils.

The Smc1 and Smc3 subunits belong to the ***s***tructural ***m***aintenance of ***c***hromosome family, representatives of which exist in eukaryotes as well as in bacteria. These proteins exist as functional dimers with homodimers and heterodimers being formed in bacteria and eukaryotes respectively. An Smc monomer consists of two globular regions, one at each of its N- and C-termini separated by an extended coiled coil-forming region (Rolef Ben-Shahar et al., 2008) (Rowland et al., 2009). In between the coiled coil sequence is placed another globular and flexible domain called the hinge, which helps the two halves of the molecule to fold back onto each other with the two extended portions together forming a 45 nm long coiled coil. The folding of the molecule brings the terminal globular regions together to form the head domain (Figure 1-1a). The arrangement gives the Smc monomers a rod-like appearance with a total length of about 65 nm while the dimers form a V-shaped structure when observed by EM. The head domains of the two SMC monomers within a cohesin molecule come together to sandwich a pair of ATP molecules in between them, imparting a ring-shaped appearance to the dimer (Figure 1-1b).

The structure of budding yeast Smc1 head homodimer in complex with an interacting stretch of the corresponding Scc1 C-terminal (Fig. 1-2. 1a and b), crystallised as a complex homodimer, has shown that the head forms a functional ATPase domain which can bind and hydrolyse ATP in a way similar to that of canonical ABC ATPases like Rad50 and the ABC transporters such as MalK and BtuCD. Like the ABC ATPase members, the Smc1 head structure is characterised by the presence of an ABC signature motif and Walker A and Walker B motifs. The ATPγS binds to the Walker A and B motifs of one monomer with its phosphate groups lying close to the signature motif of the adjacent monomer. Based on mutagenesis studies, both ATP binding and hydrolysis by Smc heads were shown to be important to generate cohesion (Arumugam et al., 2003) (Haering et al., 2004). Mutating E1158 of the conserved Walker B, required for ATP binding, results in a loss of interaction between Smc1 and Scc1 while mutating the residue S1130 of the signature motif, which compromises only the ATP hydrolysing ability of Smc heads, allows the cohesin complex to form but compromises its DNA binding ability. Thus, these studies demonstrate the ATP binding and hydrolysis of Smc heads to be important for the formation of a stable and functional cohesin core.

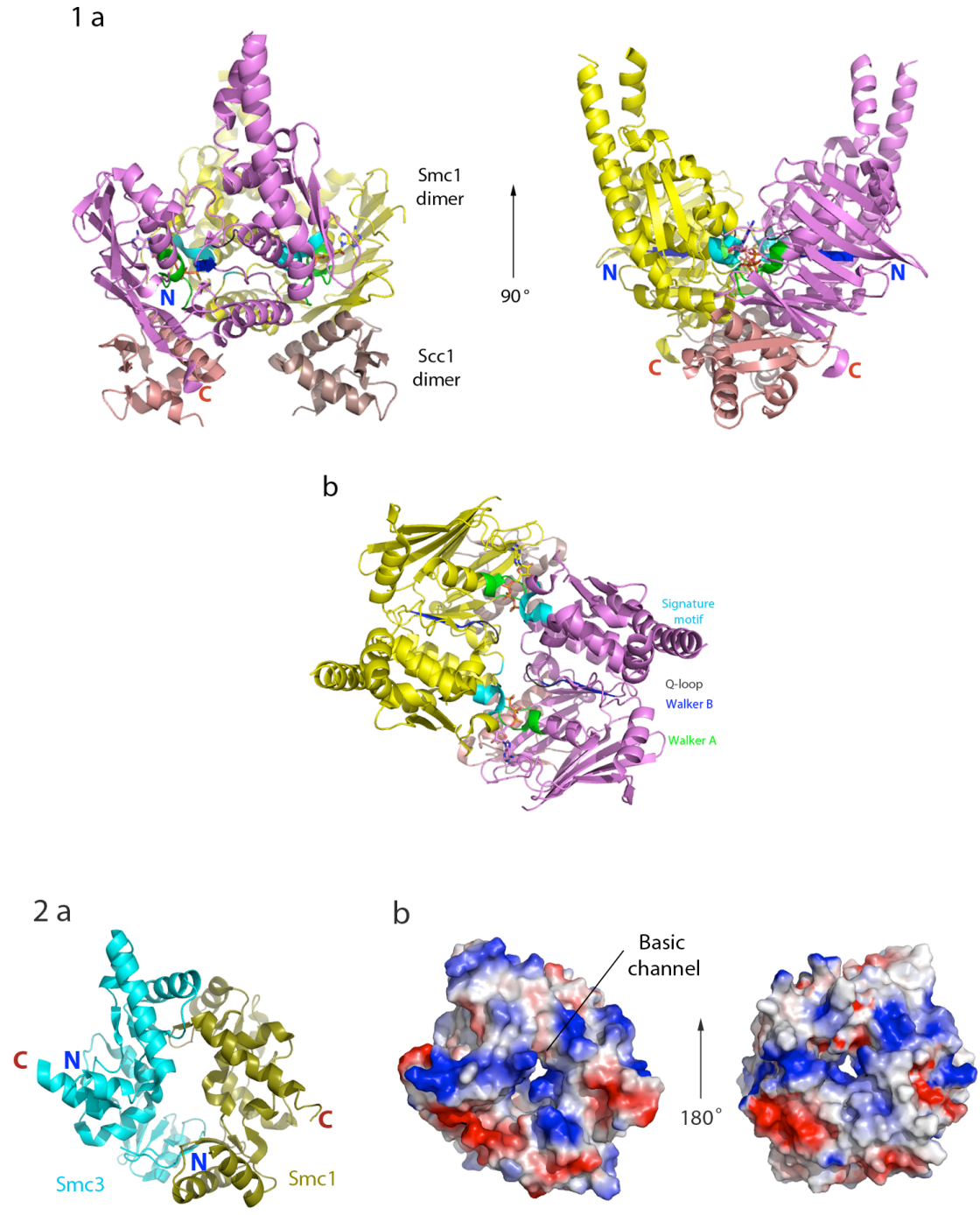


Figure 1-2: Crystal structures of cohesin subunits

(1a) Dimeric structure of Smc1 ATPase in complex with Scc1 C-terminal domain from *S. cerevisiae* and (1b) the ATPase domain as seen after rotating the structure at the top right by 90° clockwise. The Smc1 APase monomers are shown in yellow and violet and Scc1 monomers are depicted in brown. The conserved motifs are colour coded as follows: Walker A in green, Walker B in dark blue, signature motif in cyan and Q-loop in dark grey. (2a) The Smc1-Smc3 heterodimeric hinge structure from *Mus musculus* and (2b) the surface charge distribution of the dimer showing the conserved positively charged central channel

The hinge domain at the other end of the coiled-coil forms a stable dimerization interface between two Smc monomers. Crystal structures of a *Thermotoga maritime* homodimeric hinge (Haering et al., 2002) and a heterodimeric hinge from *Mus musculus* (Kurze et al., 2011) have both revealed a dough-nut shaped appearance of the dimer (Fig 1-2. 2a and b; *M. musculus* hinge structure has been shown). Each half of the doughnut shaped structure is contributed by one monomer and interacts with the other half through two binding interfaces located at opposite poles. Both the binding interfaces have been shown to be equally important in the formation of functional Smc dimers (Kabsch, 2010). The structure also revealed a conserved and positively charged narrow channel within the dimer. Charge neutralising mutations in the corresponding residues result in loss of cohesion leading to lethality although Smc dimerization and association of cohesin to chromatin of such mutants remain unaffected (Kurze et al., 2011). In addition, the intramolecular and antiparallel nature of the coiled coils emanating from the monomeric halves could also be demonstrated from these structures.

The protein Scc1 (***s***ister ***c***hromatid ***c***ohesion), also called Mcd1 (***m***itotic ***c***hromosome ***d***eterminant), is an Smc associated subunit of cohesin. In the fission yeast and higher eukaryotes, two different Scc1 isoforms exist, the mitosis specific Rad21 and the meiosis specific Rec8. In cells, the SCC1 levels fluctuate during cell cycle progression (Rowland et al., 2009) - it can be first detected during the late G1 phase but associates with chromatin during S phase and persists through G2 until anaphase onset when it starts getting degraded through APC/C activity. Scc1 is a member of the kleisin family of proteins, characterised by their association with Smc ATPase domains (Evans, 2006). Biochemical experiments have shown that binding of the N- and C-terminal domains of Scc1 to Smc3 and Smc1 ATPase heads respectively stabilises the ATP mediated head dimerisation (Figure 1-1c). This kind of arrangement of Smc1, Smc3 and Scc1 forms a tripartite ring-like structure that can mediate a stable enclosure of the sister DNA strands during cohesion. Studies have further revealed a tighter binding of Scc1 with the Smc1 head (compared to that of Smc3) and that this interaction is required for the ATPase activity of Smc1 and consequently, in the formation of a functional cohesin (Arumugam et al., 2003). Sequence analysis of the protein shows the presence of conserved N- and C-terminal domains separated by a flexible region in the middle. The

C-terminal region (last 80 amino acids in the budding yeast orthologue) has been shown to form a winged-helix domain, which interacts with the Smc1 head through conserved hydrophobic residues (Haering et al., 2004). The N-terminal of the protein, proposed to interact with the Smc3 head, also has a similar conserved region although its structure is not available and is thus not known for sure whether it can form a winged helix like its C-terminal. The middle, flexible region of the protein serves as the target for the cysteine protease separase during the metaphase to anaphase transition (Uhlmann et al., 1999; Uhlmann et al., 2000). Phosphorylation of SCC1 during mitosis triggers its cleavage at two sites, one between the residues R268 and R269 (the main site) while a second less specific site exists at R180. Separase mediated cleavage of SCC1 is an important event which ensures the release of the sister chromatids triggering their segregation during anaphase. The separase mediated cleavage of SCC1 is also considered to be an important evidence supporting the 'topological embrace' model of sister chromatid cohesion.

The other cohesin subunit is called SCC3 that associates with SCC1 to complete the cohesin core (Toth et al., 1999) (Fig. 1-1c). The protein remains associated to chromatin from late G1 phase until metaphase to anaphase transition. SCC3, like other cohesin subunits, is also conserved among eukaryotes with maximum sequence similarity in the N-terminal half. The orthologue SA (***s***tromalin ***a***ntigen) replaces SCC3 in vertebrate cohesin. In these cells, two isoforms (SA1 and SA2) of the protein exist and each can participate in forming cohesin although only one of the isoforms is enriched in a given cell type, e.g., variation of the relative abundance of the two isoforms in somatic (higher SA2 level) and egg extracts (higher SA1 level) (Losada et al., 2000) (Sumara et al., 2000).

1.3 Cohesin associated proteins

Additional proteins that associate with the cohesin core complex include Pds5, Wapl and Sororin. All these factors are characterised by their regulatory roles and influence either cohesion maintenance or removal of cohesin from DNA or both. They are also less stably bound to cohesin in comparison to the interactions among the cohesin

subunits. A detailed description of the proteins and their roles in sister cohesion is given below.

The protein Pds5 forms a sub-stoichiometric and comparatively less stable association with the cohesin complex than among the core subunits (van Heemst et al., 1999). In budding yeast, *Caenorhabditis elegans* and *Drosophila*, the protein is essential for sister chromatid cohesion as well as condensation and *PDS5* mutant strains are lethal (Hartman et al., 2000) (Panizza et al., 2000) (Dorsett et al., 2005). Higher eukaryotes contain two isoforms of the protein, Pds5A and Pds5B, both of which associate with the cohesin separately (Losada et al., 2005; Sumara et al., 2000). The two Pds5 isoforms together with the two previously described SA1 and SA2 isoforms give rise to four different forms of cohesin in vertebrates. Studies in Xenopus and human cells show less drastic effects upon depletion of the *PDS5A or PDS5B* gene with only partial cohesion loss and mitotic chromosome assembly defects being observed. Intriguingly, depletion of both *PDS5A* and *PDS5B* from Xenopus cells causes an increased concentration of cohesin in the chromosome arms although centromeric cohesion is partially weakened. Nevertheless, the protein is highly conserved among eukaryotes and a sequence analysis shows the presence of HEAT repeat sequences, known to form motifs observed in protein-protein interaction domains. Due to its predicted HEAT repeat structure, it is thought to mediate protein-protein interactions important to stabilize and maintain sister chromatid cohesion In budding yeast, Pds5 has been shown to associate with chromatin in a cohesin dependent manner and specifically relies on Scc1 for this interaction. Conversely, the stable association of Scc1 with cohesin also depends on the presence of Pds5. In addition, the protein has been shown to be important for cohesion maintenance during metaphase in budding yeast (Panizza et al., 2000).

The gene for Wapl (***w***ings ***ap***art ***l***ike) was first discovered in Drosophila. Mutation in the *wapl* gene mostly caused larval lethality while the ones that survived developed into adults with abnormally separated wings (Verni et al., 2000). Wapl has been shown to be important for regulating heterochromatin organization and chromosome segregation in the flies. The metaphase chromosomes in *wapl* mutant flies show cohesion defects in heterochomatic regions and failed to form the usual X-shaped metaphase chromosomes.

The sister chromatids in these cells lie parallel to each other due to their partial cohesion loss in heterochromatic regions. The *wapl* gene has oncogenic properties wherein its depletion in human cell lines causes cell death and overexpression of the gene in mice is tumourigenic (Oikawa et al., 2004). The human Wapl (hWapl) was shown to associate with chromatin through its interaction with cohesin and binds the subunits SCC1 and SA1/SA2 of the cohesin core (Cowtan, 2006; Kueng et al., 2006). Apart from the cohesin core, Wapl can bind both Pds5 isoforms (Pds5A and Pds5B), forming a 'subcomplex', the assembly of which depends on cohesin. In humans, the interaction with Pds5 is mediated through two conserved motifs, called FGF, located in the Wapl N-terminal, although the motif does not appear to be strictly conserved in other Wapl orthologue sequences (Shintomi and Hirano, 2009) (Figure 1-4). The protein has been shown to regulate the association of cohesin with chromatin and depletion of the gene in HeLa cells causes cohesin to persist on sister chromatids for longer periods of time causing a delayed resolution of the sisters during mitosis. Sequence database searches showed Wapl to be an evolutionarily conserved protein with putative members across eukaryotes. The identified orthologues had variable and less conserved N-terminal sequences while their C-terminal appeared to be better conserved (Figure 1-3) and were predicted to form a helical repeat domain, named as WAPL.

Based on the initial sequence analyses, a budding yeast protein, called Rad61, was found to be distantly related to the human and *Drosophila* Wapl (Kueng et al., 2006). The *RAD61* gene was identified in a genetic screen for mutants sensitive to radiations and has since been implicated to be an important determinant of cohesion and chromosome segregation. The observations that the protein Rad61 has a WAPL-like sequence in its C-terminal and associates with cohesin (through Scc1 and Scc3 subunits) and Pds5, much like the human Wapl protein, led to the conclusion that Rad61 is indeed the budding yeast orthologue of Wapl and thus, has been renamed Wpl1 (Rolef Ben-Shahar et al., 2008).

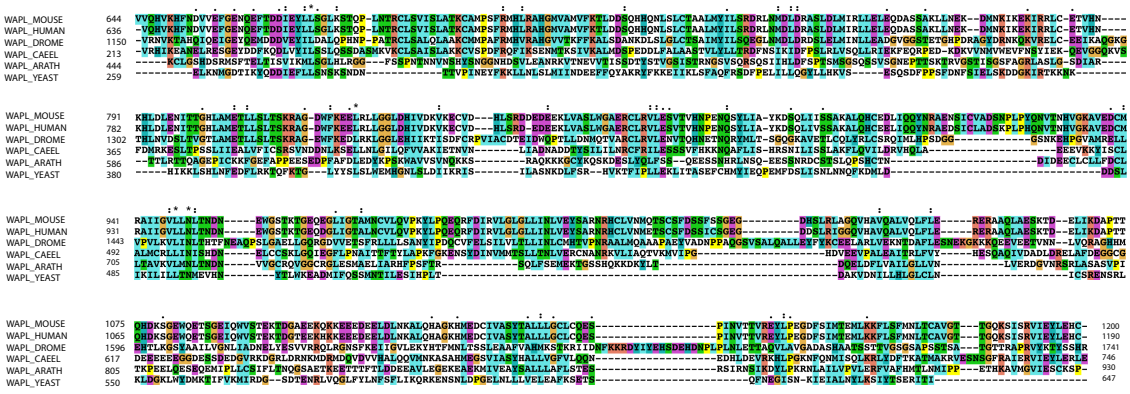


Figure 1-3: Multiple sequence alignment of the WAPL domain

The WAPL domain sequences of Mus musculus (mouse), Homo sapiens, Drosophila melanogaster, Caenorhabditis elegans, Arabidopsis thaliana and Saccharomyces cerevisiae aligned using ClustalX.

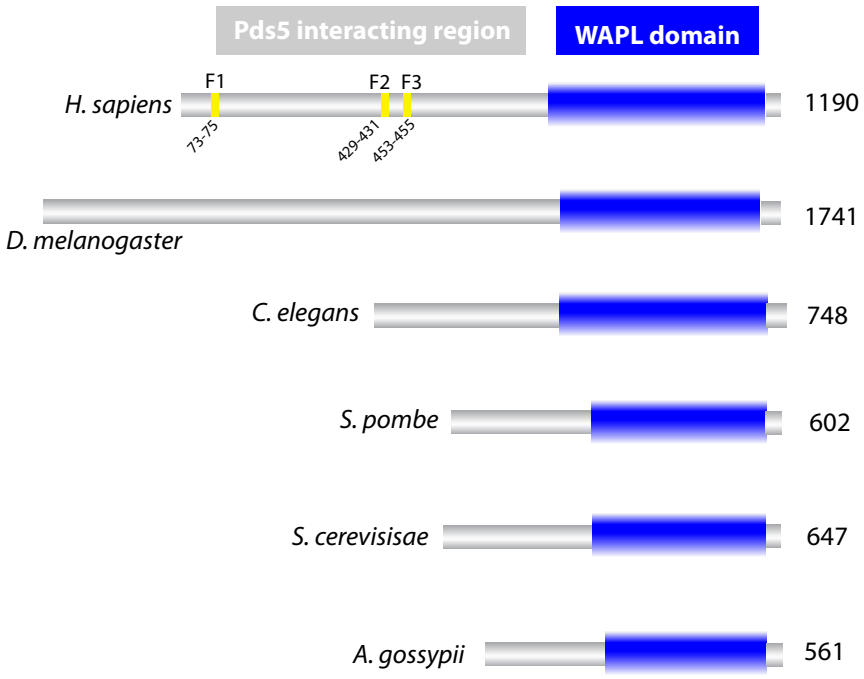


Figure 1-4: Domain organisation of WAPL orthologues

The Wapl family proteins show a variable N-terminal region (grey) responsible for interacting with Pds5. The human orthologue has been shown to contain three FGF motifs (shown as yellow bands and numbered), which are important in mediating the interaction with Pds5. The C-terminal of the protein forms a conserved domain known as WAPL (dark blue).

In spite of the striking similarities of Rad61 with its vertebrate orthologue, the roles played by the proteins in sister chromatid cohesion are apparently contradictory. Contrary to the *wapl* depletion phenotype observed in the vertebrate cells, budding yeast strains in which the *WPL1* gene is mutated show partial loss of cohesion much like the mutant flies. Based on genetic experiments in budding yeast (described in section 1.6.2), it has been proposed that Wpl1 binds simultaneously to Pds5 and Scc3 and this complex tends to maintain cohesin in a form incapable of cohesion establishment (Rowland et al., 2009). The subsequent acetylation of cohesin during S phase by Eco1 counteracts the destabilising forces of the Wpl1-mediated complex, triggering cohesion establishment. The hypothesis appears attractive and if proven, would mean that both vertebrate Wapl and the budding yeast Wpl1 mediate the same process, i.e., promoting 'anti-establishment', although the respective mechanisms may vary.

The protein sororin was initially identified as a vertebrate specific, cell-cycle regulated protein, which undergoes APC/C mediated degradation during the G1 phase. It was shown to be important for sister chromatid cohesion and normal mitotic progression (Rankin et al., 2005; Schmitz et al., 2007). The protein seemed to be less conserved as related sequences could not be identified in invertebrates. However, in a recent study, an orthologue of the protein has been shown to exist in Drosophila, called Dalmatian. Orthologues of the protein identified so far carry a C-terminal conserved domain, which has been named 'sororin'. The protein has been shown to stabilise sister chromatid cohesion in a DNA replication dependent manner while also requiring Esco1 (vertebrate orthologue of budding yeast Eco1) mediated acetylation of cohesin (Nishiyama et al., 2010). Furthermore, sororin could displace Wapl from Pds5 and the interaction was mediated through a lone conserved Wapl-like FGF motif in the N-terminal of the protein. Sororin dissociates from chromatin during prophase following its phosphorylation, which is thought to reactivate Wapl and subsequent removal of cohesin from chromosome arms through the prophase pathway.

1.4 Meiosis specific events and subunits

During meiosis, DNA replication is followed by chromosome segregation in two steps, called meiosis I and meiosis II, a process important in giving rise to haploid gametes from diploid parental cells. During prophase I, a pair of homologous chromosomes (one paternal and one maternal) each of which consists of two sisters, need to be aligned opposite to each other. The chromosomes arranged in this fashion, termed as 'bivalents', are required for chiasma formation between non-sister chromatids and thus, for proper homologous recombination. As a consequence of bivalent formation, the sister kinetochores are captured by the spindle apparatus emanating from the same pole during the subsequent metaphase I stage. This situation is different from mitosis where the sister kinetochores are captured by spindles from opposite poles. During metaphase I to anaphase I transition, cohesin from the arms of sister chromatids are removed and this helps in the separation of the homologous chromosomes causing the sister chromatids to move together to opposite poles during anaphase I. The second meiotic division is characterized by the removal of centromeric cohesin which ultimately releases the sister chromatids during anaphase II.

Meiosis specific cohesin subunits exist that allow for the specific events to occur during the process. One such subunit and also the most important one is Rec8, which is a homologue of the kleisin Scc1 of budding yeast and Rad21 of higher eukaryotic mitotic cohesin (Klein et al., 1999). The protein derived its name based on its higher homology to the meiosis specific kleisin Rec8 rather than the mitosis specific Rad21 of *S. pombe*. In budding yeast, the Rec8 protein has 16% identity and 34% similarity to Scc1. Rec8 has been shown to be important in the formation of the synaptonemal complex and axial filament and is also essential for meiotic homologous recombination. The protein specifically forms meiotic cohesin and is found to localise with cohesin on chromatin starting from premeiotic DNA replication, its levels decreasing drastically during meiosis I and is removed completely from chromosomes in anaphase II. All the meiosis specific events that require Rec8 are also aided by the Smc3 subunit of cohesin although unlike Rec8, only one form of Smc3 is involved both in mitosis and meiosis. Like Scc1, Rec8 is removed from the cohesin arms during prophase of meiosis I while being

protected at the centromeres by the combined activity of shugoshin-PP2A complex (1994; Murshudov et al., 1997). However, it has been shown that this protection during meiosis is mediated by antagonising the effect of the kinase CK1 (casein kinase 1δ/ε) rather than Plk1 (involved in mitosis) by the shugoshin-PP2A pair (Ishiguro et al., 2010). An important process during meiosis is the co-orientation of sister kinetochores during meiosis I. A protein known as monopolin has been implicated to play a key role in this process by promoting syntelic attachment of sister kinetochores during budding yeast meiosis I (Petronczki et al., 2006). Monopolin probably interacts with specific kinetochore subunits, including Dsn1 of MIND/Mis12 and the Mif2/CENP-C. It has been proposed that the protein acts like a clamp that cross-links the two adjacent sister kinetochores, preventing their amphitelic attachment.

1.5 General features of ABC ATPase domains

The ABC (***A***TP-***b***inding ***c***assette) ATPase family proteins are characterised by the ability to bind and hydrolyse ATP, the energy derived from which is coupled to their physiological activities. The ABC ATPase domain, also called the nucleotide-binding domain (NBD), generally form functional dimers by cooperatively sandwiching a pair of ATP molecules in between the monomers. Initial understanding of the mechanism of functioning of these proteins were derived from the crystal structures of the ATPase domains of membrane transport proteins, e.g., HisP (Hung et al., 1998), MalK (Chen et al., 2003) and BtuCD (Locher et al., 2002) and the double-strand break repair protein Rad50 (Hopfner et al., 2000), all of which share conserved motifs involved in either ATP binding and hydrolysis, dimerisation or cross-talk with other domains. The Smc proteins were predicted to harbour similar ATPase domains and later, based on the respective crystal structures, the NBDs of these indeed showed an ABC ATPase-like domain organisation. A detailed description of the ABC ATPase domains of membrane transport proteins and Rad50 is given below to provide a general understanding of their functioning.

1.5.1 ABC transporters

The ABC transporters are dimeric proteins generally consisting of two types of subunits – a transmembrane (TM) subunit and a hydrophilic, non-membrane bound subunit. The TM subunit consists of two identical membrane-embedded domains, which together form a gate for the molecule to be transported while a pair of the ABC ATPase domains forms the hydrophilic subunit. The ABC ATPase domain is characterised by the presence of a series of conserved motifs involved mainly in ATP binding and hydrolysis either directly or through water molecules. Some of these conserved motifs also create a connection with the membrane-spanning region, thought to be important for the transport process.

Each monomeric NBD consists of a six-stranded β-sheet forming the core, which is surrounded by about nine α-helices and an additional three-stranded β-sheet. These may be divided into two subdomains – a RecA-like subdomain consisting of two β-sheets and six α-helices and a smaller subdomain with three to four helices (Figure 1-5a). An ATP is sandwiched in between the RecA subdomain of one monomer and the helical subdomain from the adjacent monomer. Among the most conserved of the motifs characterising the domain are the Walker A (also called P loop) and Walker B motifs in the RecA-like subdomain and the signature motif (also called linker peptide, LSGGQ or C-motif) that is part of the helical subdomain. The Walker A motif consists of a loop and its residues make contacts with all the three phosphates of the ATP while the Walker B motif, which forms a β-strand, coordinates the Mg^{2+} ion through a water molecule bound by its conserved aspartate and also through a glutamate positioned immediately after the motif. The glutamate is required for the catalytic activity of the protein as mutating the residue to glutamine locks the ATPase monomers in a dimeric state by preventing ATP hydrolysis. The Walker A and B motifs of one monomer sandwich an ATP with the help of a signature motif of the other monomer. The signature motif contacts the γ-phosphate and its conserved serine helps in hydrolysing the ATP. Preceding residues of both Walker A and the signature motif make further contacts with the nitrogen base as well as the ribose of an ATP molecule.

The ATPase domain is characterised by additional motifs, which are the Q loop, the H motif and the D loop. The conserved Q loop forms a flexible linker between the RecA-like domain to the helical domain and a conserved glutamine residue in the motif forms a contact with the Mg^{2+} and the attacking water molecule. The Q loop is thought to be important to link the hydrolysis of ATP to conformational changes in the transmembrane domain that facilitates transport. This is aided by the flexibility of the motif and also by its proximity to the catalytic site. Importantly, the loop has been shown to interact with two angled helices (L loop) of the TM domain and thus might form a connecting link to transmit information from the ATPase to the TM domain. The H-motif is characterised by the presence of a conserved histidine which hydrogen bonds with the γ-phosphate of the ATP while the D loop forms dimer contacts between the two NBD monomers.

The hydrolysis of ATP is thought to bring about conformational changes, which facilitate transport by the ABC ATPase transporters. The evidence comes from the differences in the relative orientations of the helical domain with respect to the RecA-like domain in the ATP free and the ATP bound structures. In ATP-free or AMP:PNP bound states (both lack the γ-phosphate), the helical domain assumes different relative orientations in different structures. However, the domain loses its flexibility and is locked into the same orientation in all the structures in presence of ATP. This might be important for the helical domain to regain its flexibility once ATP is hydrolysed to ADP, which lacks the γ-phosphate. In this situation, the movement of the domain might result in retraction of the signature motif in order to release the ADP. In addition, the flexibility might result in the ATPase activity being controlled by the TM region by virtue of its ability cross-talk with the ATPase domain.

1.5.2 The Rad50 ATPase

Rad50 is a DNA double strand break repair protein required for homologous recombination and non-homologous end joining (NHEJ) repair pathways. The protein forms a functional complex with two other proteins - Mre11 and Nbs1, together known as the MRN complex. This complex has also been implicated in other related pathways

including telomere maintenance and DNA damage checkpoint control. Rad50 can both bind and mediate partial unwinding of double-stranded DNA in an ATP-dependent manner.

Much like the Smc proteins, Rad50 also contains terminal globular regions at either end separated by an extended sequence (600-900 residues long). Two such Rad50 monomers dimerise in an antiparallel, head-to-tail arrangement to form catalytic halves at each end. However, unlike the Smc proteins, which form ring shaped functional heterodimers, the Rad50 molecules form open or linear structures, made of two dimers (dimer of dimers) linked at their globular termini, thus resembling a double dumb-bell. The interacting double dumbbell globular heads form functional ABC ATPase domains. The Rad50 ATPase provided structural and mechanistic insights into various then unexplained features such as the significance of the conserved ABC signature motif, the mechanism of cooperativity of ATP binding and hydrolysis and significance of the ATP-induced conformational changes. Furthermore, the structure also provided an explanation for the diseased phenotypes linked to specific mutations in both Rad50 (*rad50S* phenotype in yeast) as well as in the CFTR (cystic fibrosis transmembrane regulator; another ABC ATPase member) protein.

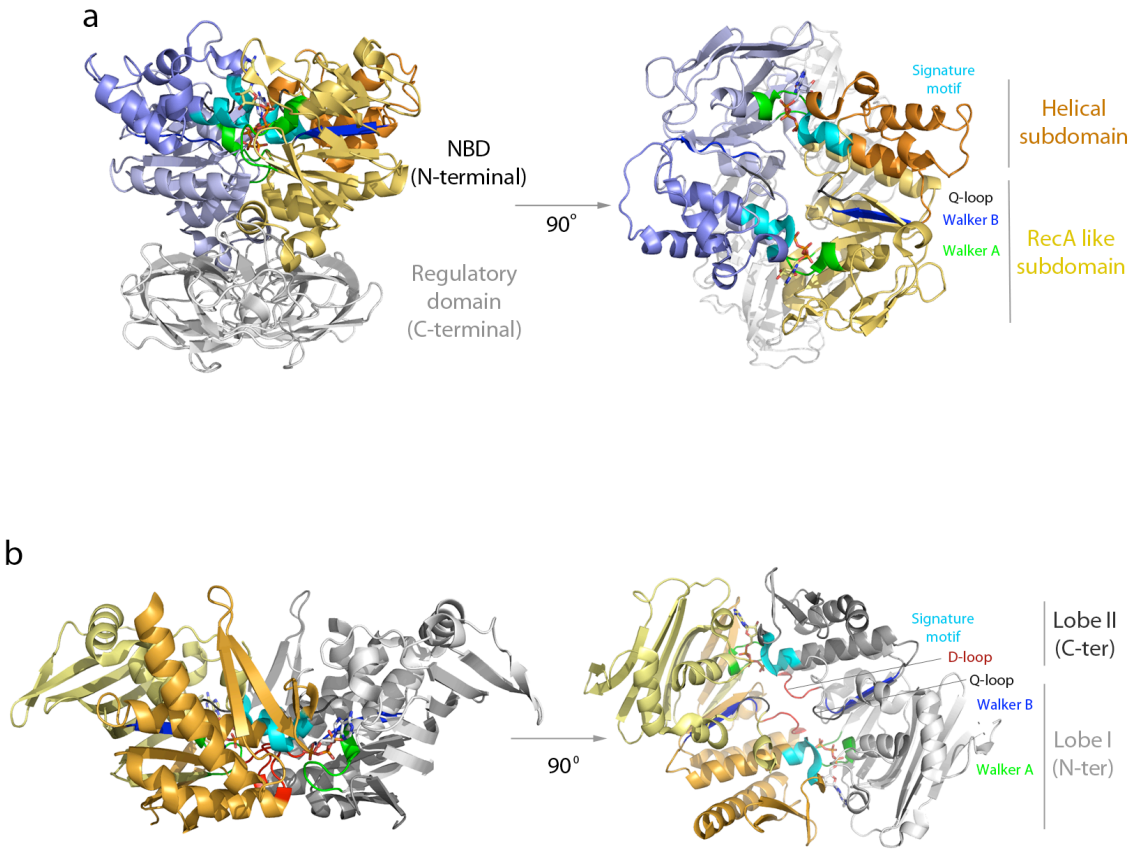


Figure 1-5: ABC ATPase architecture

The ABC ATPase catalytic dimers formed by (a) the E. coli maltose transporter (MalK) and (b) Rad50 from *Pyrococcus furiosus*. The domain organisation as well as the most conserved and important motifs in each structure has been depicted (Walker A in green, Walker B in dark blue, signature motif in cyan, Q-loop in dark grey and D-loop in red)

The Rad50 protein (from *Pyrococcus furiosus*) was crystallised both in ATP-unbound and bound states. The ATP unbound form showed the structure assumed by one half of the catalytic domain formed by the assembly of the N-terminal of one monomer with the C- terminal of the other, together forming an ellipsoidal and bilobed domain. The larger Lobe I, made up by the N-terminal of the first monomer, forms an α/β roll while the smaller Lobe II, made up of the C-terminal of the second monomer, is organised as a β-α-β sandwich.

The ATP bound form showed a double-dimeric structure (dimer of dimers) and formed a functional ATPase catalytic domain (Rad50cd) (Figure 1-5b). The Rad50cd assumed a

disc-shaped structure with a concave and a convex surface. The dimeric interface forms a shallow groove with a positively charged concave surface extending along the entire interface, which was proposed to be the DNA binding site of the ATPase. Mutations previously observed in the yeast *rad50S* cells, deficient in meiosis, could be mapped mostly onto a surface formed by the β-sheet of Lobe I that is adjacent to the DNA binding groove. The mutations were proposed to affect binding of a meiosis specific factor.

The individual dimers of the Rad50cd were joined together with the help of a pair of sandwiched ATP (without Mg^{2+}; so non-hydrolysable) or AMP:PNP molecules. Each ATP was sandwiched in between the Walker A motif of one monomer and the signature motif of the other. Further, it was seen that the γ-phosphate O of the ATP bound to the signature motif conserved residues serine (S793; through side chain O) and glycine (G795; through main chain N). Mutation of this serine to an arginine resulted in the constitutive association of the dimeric heads in the presence of ATP, resulting from a loss in the catalytic ability of the mutant proteins. This showed the involvement of the signature motif both in ATP binding as well as hydrolysis and thus the strict conservation of this stretch throughout the ABC ATPase superfamily. The observation also showed the basis of the cystic fibrosis phenotype observed when the S549 residue in the CFTR signature motif (based on sequence alignment) is mutated. Comparison of the ATP free and bound structures revealed a 9Å shift in the position of Lobe II relative to Lobe I caused by ATP binding/hydrolysis. This observation together with the finding that the Walker A and the signature motifs, each located in two different monomers, are needed together to bind as well as hydrolyse the ATP molecules provided the basis for the cooperativity of catalysis by the Rad50cd.

1.6 The cohesion cycle

Sister chromatid cohesion is a cyclic process, which requires the cohesin complex to be first loaded during the late G1 phase/telophase. Cohesion is subsequently 'established' during S phase and maintained until metaphase-to-anaphase transition (Uhlmann and Nasmyth, 1998). Once biorientation is achieved during metaphase, the role of cohesin in sister cohesion is fulfilled and the chromosomes now need to give-in to the outward spindle-pulling forces. The cohesin rings at this stage, are removed from the chromosomes, resulting in setting the sisters free to move toward the opposite poles during anaphase and the event marks the completion of a complete cohesion cycle. A description of the individual steps of the cohesion cycle, i.e., cohesin loading, establishment and dissolution of cohesion, have been given below while explaining the specific factors needed at each stage.

1.6.1 Cohesin loading and binding to DNA

Each cycle of cohesion starts with the association of the cohesin rings with chromatin during G1 phase in budding yeast (Guacci et al., 1997b) (Rowland et al., 2009) and telophase in vertebrate cells (Losada et al., 2000; Sumara et al., 2000). The process, known as cohesin loading, is brought about by the recruitment of a complex formed by the conserved proteins Scc2 and Scc4 (Furuya et al., 1998) (Ciosk et al., 2000) (Gillespie and Hirano, 2004) (Bernard et al., 2006). Orthologues of both Scc2 and Scc4 have been identified in other organisms and the fission yeast Mis4, Drosophila Nipped-B and the human Nipped-B like (NIPBL) are the Scc2 counterparts in the respective organisms. Mutations in the human orthologue, NIPBL, cause the multisystem developmental disorder Cornelia de Lange Syndrome (Krantz et al., 2004).

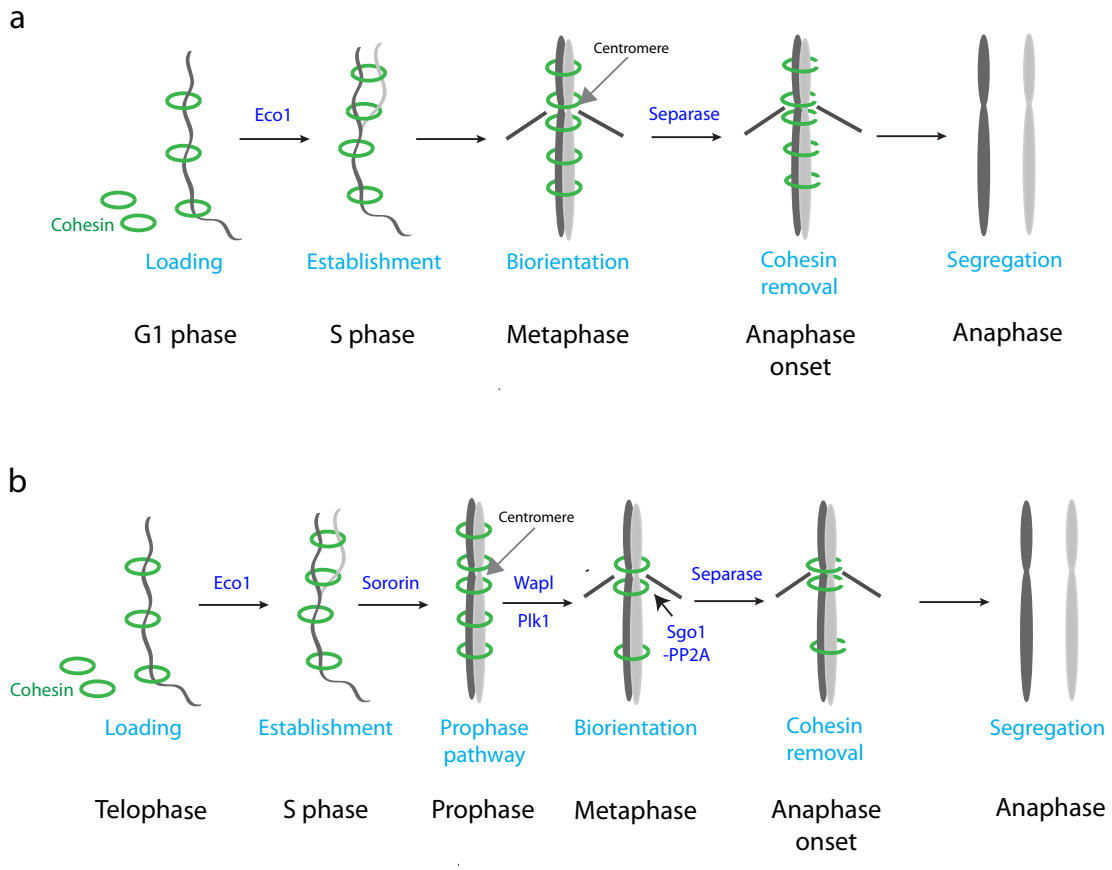


Figure 1-6: The cohesion cycle

The cohesion cycle in (a) budding yeast and (b) vertebrates. Loading of cohesin occurs in either G1 phase (in budding yeast) or telophase (in vertebrate cells). Cohesin removal in budding yeast is a single step process, which occurs during metaphase-to-anaphase-transition. In vertebrate cells cohesin is removed in two steps – initially from the chromosome arms during prophase and subsequently from the centromeres during metaphase-to-anaphase transition.

Scc2 was identified in the initial genetic screens for cohesion-defective budding yeast mutants. The protein was found to associate with the cohesin core in a non-stoichiometric ratio although its interaction with Scc4 was stoichiometric. Mutations in either *SCC2 or SCC4* genes cause a premature separation of sister chromatids, much like other cohesin core subunit mutants. However, unlike cohesin, the Scc2/Scc4 complex is required only until DNA replication during S phase and is dispensable for cohesion maintenance during G2 and metaphase showing the importance of the complex specifically for cohesin loading. The complex can stably associate with chromatin and the individual components depend on each other for this interaction. Scc2 consists of a conserved C-terminal predicted to form HEAT repeats while a variable N-terminal region exists in the vertebrate orthologues. The N-terminal of

vertebrate Scc2 has been shown to mediate its interaction with Scc4 and the complex formed between Scc4 and a truncated Scc2 consisting of the N-terminal only is sufficient to bind chromatin. However, this interaction does not seem to be enough to load cohesin onto chromatin suggesting a role for the conserved C-terminal of Scc2 in the process (Takahashi et al., 2008).

Although the importance of the Scc2/Scc4 complex in cohesin loading is well established, the exact molecular mechanism of this reaction is still not clear. ATP hydrolysis by the Smc heads is thought to be one of the important factors mediating the process. Smc head mutants incapable of either ATP binding or hydrolysis have been shown to be defective in cohesion (Arumugam et al., 2003) (Weitzer et al., 2003). ATP hydrolysis could give rise to transiently opened cohesin rings through which DNA can gain entry. It has been further shown that cohesin incapable of ATP hydrolysis can only undergo attachment with chromatin but cannot entrap DNA. Entrapment of DNA by cohesin and its subsequent sliding has been shown to be dependent on ATP hydrolysis (Hu et al., 2011).

Apart from the ATPase activity of the head, the hinge is also considered to have a role in stabilising the association of cohesin with chromatin (Gruber et al., 2006; Hirano and Hirano, 2002). Artificial inhibition of hinge opening prevents cohesin to associate with chromosomes and establishment of cohesion. In addition, mutations aimed to disrupt or weaken hinge mediated dimerisation also made cohesin incapable of stably interacting with DNA (Kabsch, 2010). As suggested by some of the studies, an interaction between the head and the hinge, either direct or indirect, might be required for a stable interaction between cohesin and chromatin (Mc Intyre et al., 2007).

Once cohesin is loaded onto DNA, the complex generally gets enriched at specific sites and this distribution is thought to be crucial for regulating gene expression. The cohesin attachment sites on DNA have been studied using ChIP (chromatin immunoprecipitation) experiments. In budding yeast, the sites of attachment of cohesin on chromosome arms are called cohesin attachment regions (CARs) and are characterised by AT rich sequences (Blat and Kleckner, 1999). CARs are found at the

sites of convergent transcription and are made up of 0.8 kb regions located at about 10 kb intervals (Lengronne et al., 2004). These sites of cohesin association are different from the sites of cohesin loading which is thought to be the sites where Scc2 binds DNA. Thus, cohesin is thought to relocate from the sites of their loading to transcription end sites and this process has been proposed to occur passively due to the movement of the RNA polymerase II transcription complex. Differences have been observed between the CARs in budding yeast and the cohesin binding sites in higher eukaryotes. In *Drosophila*, cohesin binds to transcribed regions, which has been suggested to be facilitated by transcription-mediated chromatin unwinding (Misulovin et al., 2008). The observation that Nipped-B and cohesin bind similar regions on *Drosophila* non-repetitive sequences is also in contrast to the situation in budding yeast. In humans, majority of cohesin is found in intergenic regions and at introns while a smaller subset localizes at sites directly upstream or downstream of genes. In addition, human and mouse cohesin is also found at the binding sites of CTCF insulators although these do not correspond to Scc2 binding sites and depletion of cohesin recapitulates the effects of CTCF loss (Wendt et al., 2008).

Interaction of cohesin with DNA and its localisation at specific sites has been implicated in processes as important as gene transcription and expression. However, deciphering the role of cohesin in these processes depends on insights into the underlying mechanism of cohesin movement and its organism-specific localisation on the DNA, aspects that are not yet well-understood.

1.6.2 Establishment of sister chromatid cohesion

The cohesin complex can associate with chromatin throughout interphase, however, sister chromatid cohesion can only occur if an intact and stable cohesin complex is formed at the latest by S phase (Uhlmann and Nasmyth, 1998). Accordingly, budding yeast cells lacking Scc1 during S phase cannot align and segregate their chromosomes during mitosis even if Scc1 expression is switched on starting from G2 phase. It was further shown that lethality caused due to the lack of Scc1 could be overcome by blocking DNA replication. This initial study showed that loading of cohesin onto DNA

is not by itself sufficient to form cohesive sisters but requires an additional step, called 'establishment', which is linked to DNA replication. It was subsequently found that a protein called Eco1 (***e***stablishment of ***co***hesion), also called Ctf7 (***c***hromosome ***t***ransmission ***f***idelity) is an essential factor required for the establishment of cohesion during S phase (Williams et al., 2011) (Toth et al., 1999). *ECO1* mutant cells showed precocious sister chromatid separation and unlike the cohesin subunits, its presence during S phase (rather than throughout interphase) is sufficient for proper segregation of chromosomes. In addition, the mutant phenotype of these cells was overcome by overexpression of the *PCNA* gene. These experiments not only reiterated the requirement of a separate establishment step but also identified an important factor necessary for this DNA replication dependent reaction.

A functional link between establishment of cohesion and DNA replication became more apparent after Eco1 was shown to interact physically with several proteins that assemble at the replication fork. Firstly, Eco1 was shown to associate with all the three different RFC complexes, which have non-overlapping roles during replication (Kenna and Skibbens, 2003). Second, Eco1 along with Ctf4 and Ctf18, proteins that have also been shown to be required for cohesion, were found to localise at the replication fork and Ctf4 was actually found to move with the fork along chromosomes (Lengronne et al., 2004). Finally, PCNA was also shown to physically interact with Eco1 in budding yeast and the human orthologue Esco2 (Moldovan et al., 2006). This interaction was shown to be required for the establishment of cohesion while additional genetic experiments suggested that cohesion might be suppressed by sumoylation of PCNA, again showing the link between replication and cohesion.

Eco1 sequence analysis shows a conserved acetyltransferase domain on the C-terminal with an acetyl-coenzyme A (Acetyl CoA) binding site similar to that of the GNAT (Gcn5-related N-acetyltransferase) family (Ivanov et al., 2002). Eco1 was shown to acetylate the proteins Scc1, Scc3 and Pds5 *in vitro* but not histones showing its specificity for cohesin subunits. However, acetylation could not be detected in the proteins (Scc1, Scc3, Pds5) upon immunoprecipitation from yeast cells, showing that these subunits may not be the *in vivo* acetylation targets of Eco1. The N-terminal of the

protein has a C_2H_2–like zinc-finger motif (residues 33-57 in the budding yeast orthologue), which differ from similar motifs in other proteins by an insertion of two additional residues within a loop in the centre. The zinc-finger motif has been shown to be important for the functioning of the protein and enhances the efficiency by which the enzyme acetylates its substrates (Leslie, 2006).

The *in vivo* target of Eco1 was subsequently shown to be the Smc3 subunit of cohesin using a combination of genetic and biochemical studies in budding yeast and humans (Rolef Ben-Shahar et al., 2008) (Unal et al., 2008) (Zhang et al., 2011). It was observed in budding yeast that a spontaneous mutation of a conserved lysine K113 to an asparagine in Smc3 could suppress the lethality of a temperature sensitive *ECO1* mutant (*eco1-1* strain). Immunoprecipitation experiments showed that this lysine (K113) along with the preceding one (K112) are acetylated in the wild type cells *in vivo* in an Eco1 dependent manner and since asparagine is a known mimic of acetylated lysine, the observation provided an explanation for the lysine to asparagine mutation in the suppressor mutants. If the conserved lysines are mutated to arginines, which are non-acetylatable but have similar side chain chemistry as lysine, then the mutant cells show severe cohesion defects (in case of K113R single mutants) or are not viable at all (K112/113R double mutants). Acetylation of Smc3 was further shown to be cell cycle dependent and is maximum during the S phase and correlates exactly with the timing of cohesion establishment. These results showed acetylation of Smc3 by Eco1 during S phase to be an important event that leads to the establishment of cohesion.

An additional group of *eco1-1* suppressors were identified which had nonsense mutations within the *WPL1* gene (giving rise to a truncated/non-functional gene product). In accordance with this, deletion of *WPL1* from either the *eco1-1* temperature sensitive mutant or an *ECO1* deleted strain (lethal in both cases) restores viability of the cells. Also, the lethality caused by the non-acetylatable K112R, K113R mutations in Smc3, can be overcome by deletion of *WPL1,* as in this case the cells regain viability. Based on these results it was concluded that in budding yeast, Eco1, through its acetylation, helps to overcome a destabilising effect of Wpl1 on cohesin. As a further extension of the above observations, additional suppressors of the *eco1-1* mutant were

isolated which mapped to the Pds5 and Scc3 genes and led to the idea that Wpl1 together with Pds5 and Scc3 forms an 'antiestablishment' complex which together act on cohesin to maintain it in a destabilised form (Rowland et al., 2009). Consistent with this proposal, it was shown with purified budding yeast proteins that Wpl1, Pds5 and Scc3 could indeed form a complex within themselves.

Smc3 acetylation during S phase helps in the establishment of cohesion by opposing the destabilising forces on cohesin. Acetylation is maintained through G2 until anaphase onset and then it becomes undetectable. The histone deacetylase Hos1 has been shown to be required for the deacetylation of Smc3 during anaphase and this process is dependent on cohesin dissociation from chromosomes (Borges et al., 2010). Importantly, only the deacetylated form of cohesin can act as a substrate for the next round of loading/establishment reactions and cells in which Smc3 is constitutively acetylated (either by *HOS1* deletion or by expressing Smc3 K112N K113N - the acetyl-mimic Smc3 mutant) show prominent cohesion defects.

Apart from the normal S phase cohesin activation, cohesion can also be established during G2 phase in cells carrying double-strand breaks (DSB). The cohesin subunits Rad21 of *S. pombe* and Wpl1/Rad61 were initially identified while screening for mutants defective in DNA repair giving rise to radiation sensitive phenotype (Birkenbihl and Subramani, 1992) (Game et al., 2003). Subsequently, cohesin has been implicated to also aid in the repair of damaged DNA apart from its regular function in sister chromatid cohesion. It was found that sister chromatid cohesion is a prerequisite for DSB repair during G2 phase and that the *eco1-1* mutants which were defective in cohesion establishment were also inefficient in this post-replicative repair pathway (Strom et al., 2007) (Unal et al., 2007). In the presence of DSBs, cohesin gets accumulated at a 16kb long domain around the lesion in addition to cohesin loaded at CARs and this additional enrichment is triggered by the phosphorylation of the histone H2AX specifically in this region (Unal et al., 2004). These distinct cohesin populations together become cohesive in response to DSB during the G2 phase. Generation of DSB induced cohesion, like that in S phase, requires the acetyltransferase activity of Eco1 but unlike in normal cells, this reaction was shown to be independent of replication.

Also, overexpression of Eco1 during G2 phase was shown to trigger cohesion in cells free of DSBs. Furthermore, breaks in a particular chromosome simultaneously triggered a genome wide cohesion, even at regions or chromosomes without any lesion and this further showed the replication-independent nature of the process.

Although cohesion during G2 phase requires Eco1 activity, it has been shown to involve a mechanism distinct from that of S phase cohesion. A break in the DNA is thought to trigger the checkpoint kinase Mec1/ATR that in turn activates the Chk1 kinase. Activated Chk1 was shown to be responsible for phosphorylating a conserved serine S83 of Scc1. Mutating this serine to alanine does not support cohesion in G2 phase even in the presence of DSBs whereas mutating the residue to an aspartate, considered to be a phospho-mimic mutant, allows generation of cohesion during G2 even in the absence of DSBs. Interestingly, the conserved serine is mostly specific to the Scc1 orthologues in eukaryotes but not in its meiotic homologue Rec8. In the budding yeast Rec8 sequence, an asparagine (N93) is found at a similar position to S83 of Scc1 and mutating this residue to a serine or a phospho-mimic aspartate allows generation of cohesion during the G2 phase in cells expressing the mutant forms of Rec8 (either Rec8-N93S or Rec8-N93D). In humans, however, Rec8 contains a conserved serine at a similar position instead of the Scc1 orthologue and thus, Rec8 and not Scc1 can mediate DSB-induced cohesion during G2. In a further study, it has been proposed that phosphorylation of the residue S83 in Scc1 triggers its Eco1 mediated acetylation in cells with DSB during G2 phase (Unal et al., 2007). The residues K84 and K210 in Scc1 have been proposed to be the acetylation targets of Eco1 specifically during G2 as mutating these to non-acetylatable arginines (K84R, K210R) makes cells incompetent of DSB-induced cohesion while changing them to acetyl-mimic glutamines (K84Q, K210Q) can induce cohesion even in the absence of Eco1. It was also shown that the effect of the K84R, K210R mutations in Scc1 recapitulated the effects of expressing the non-phosphorylatable S83A mutant of Scc1, that is, cohesin in these cases can associate with chromosomes around the DSB and also at CARs but cohesion cannot be established during G2. However, in this study acetylation of Scc1 could not be detected in the cells and thus, the hypothesis that the phosphorylation event regulates its proposed acetylation awaits *in vivo* confirmation.

The Scc1 subunit has also been shown to undergo sumoylation in response to DSBs and this modification is dependent on the E3 ligase Nse2 that is a subunit of the repair specific Smc5/6 complex (McAleenan et al., 2012) (Almedawar et al., 2012). Sumoylation of Scc1 does not require it to be phosphorylated and occurs along the length of the protein but is concentrated more on the C-terminal region. This process has also been shown to be required for DSB-dependent cohesion as cells expressing a sumoylation deficient Scc1 mutant show cohesion loss during the G2 phase .

Another protein that has been shown to play a major role in establishment and maintenance of cohesion is the vertebrate specific sororin. The protein has been shown to stabilise the interaction between cohesin and DNA and functions in a replication dependent manner while also relying on Smc3 acetylation. The protein was further shown to displace Wapl from Pds5, which helps to overcome the destabilising effect of Wapl on cohesin (Nishiyama et al., 2010).

1.6.3 Removal of cohesin

1.6.3.1 *The Prophase pathway*

As briefly discussed earlier, in vertebates, cohesin is removed from chromosomes in a stepwise manner – cohesin from chromosome arms is first removed during prophase by a process termed as the prophase pathway whereas the remaining cohesin molecules which persist at the centromeres are removed during the metaphase-to-anaphase transition by a distinct pathway (Waizenegger et al., 2000). During vertebrate prophase, the cohesin-associated proteins Scc2/Scc4, Wapl, Pds5A, and sororin are removed from the chromosomes. In mitotic cells, it has been shown that the cohesin subunits Scc1 and SA1/SA2 are phosphorylated during prophase in a Plk1 (polo-like kinase) dependent manner (Losada et al., 2002) (Sumara et al., 2002). Depletion of Plk1 from *Xenopus* cells caused increased association of cohesin with chromosomes whereas expression of a recombinant Plk1 in these cells restored cohesin dissociation during mitosis. Furthermore, expression of a phosphorylated form of SA2 in the Plk1 depleted cells resulted in dissociation of cohesin even from interphase chromosomes. These

experiments showed that Plk1 mediated phosphorylation of the cohesin subunits Scc1 and SA2 is an important mechanism to modulate interaction of cohesin with chromatin and is required for cohesin dissociation during prophase. Consistent with this observation it was found that expression of a non-phosphorylatable mutant of SA2 but not Scc1 results in a reduction in cohesin dissociation from chromosomes during prophase. This also showed phosphorylation of SA2 rather than Scc1 is crucial. It has been proposed that Plk1 is recruited to cohesin on the chromosome arms by sororin (Zhang et al., 2011). Using *in vitro* experiments, sororin was shown to be phosphorylated in a Cdk1 dependent manner and phosphorylation at one of these sites, a conserved motif, $ST^{159}P$, was shown to be important for its interaction with Plk1. Sororin and Plk1 were found to colocalise on chromosomes and cells expressing a T159A mutant form of sororin were defective in resolving arm cohesion.

The mitotic kinase Aurora B is also important for cohesin dissociation during prophase although it has not been shown to phosphorylate any cohesin subunit *in vitro* suggesting an indirect mechanism of regulation (Losada et al., 2002). Condensin I and Sgo1 binding to chromosomes is important for cohesin dissociation and importantly, both of these proteins were shown to be controlled by Aurora B. Condensin I, however, binds to chromosomes only in prometaphase and thus its exact role in cohesin dissociation during prophase is not yet well understood (Lipp et al., 2007).

The phosphatase Ssu72 is a conserved eukaryotic protein and has been shown to be another important regulator of cohesin dissociation during prophase (Kim et al., 2010). The human orthologue (HsSsu72) interacts both with Rad21 and SA2 directly *in vitro* and *in vivo* and was shown to localise with chromatin bound cohesin during interphase and prophase but dissociates from chromosomes in metaphase. Overexpression of HsSsu72 was shown to decrease the dissociation of cohesin from chromosomes while its depletion enhanced cohesin dissociation. It was also observed that overexpression of the protein resulted in an increase in the percentage of chromosomes with closed arms whereas its depletion caused an increase in the number of open arms, thus showing the importance of this phosphatase in only arm cohesion. Depletion of this phosphatase resulted in hyperphosphorylation of the SA2 subunit and it was shown using

recombinant proteins *in vitro* and also *in vivo* experiments that HsSsu72 causes dephosphorylation of specifically the SA2 subunit but not Rad21. It was further shown that Ssu72 regulates cohesin dissociation in a Wapl independent manner. Accordingly, in Ssu72 knocked down cells, depletion of Wapl (which by itself causes increased association of cohesin with chromosomes; described below) could not rescue the enhanced cohesin dissociation. However, in Wapl depleted cells, knocking down Ssu72 was shown to counteract the enhanced cohesin association with chromosomes. These observations show that the phosphatase activity of HsSsu72 is required to prevent hyperphosphorylation of SA2, which in turn helps to keep cohesin stably bound to the chromosome arms and thus, the process has been proposed to be an important mechanism for maintenance of arm cohesion.

The prophase pathway of cohesin removal also depends on the protein Wapl, which acts independently of phosphorylation by Plk1. Wapl depletion, in fact, has been shown to cause more severe defects in arm resolution compared to inactivation of Plk1, Aurora B or condensin I (Cowtan, 2006) (Kueng et al., 2006). Wapl is required for normal progression during mitosis as Wapl depletion causes a delayed anaphase onset due to an increased residence of cohesin on the mitotic chromosomes. Overexpression of Wapl, on the contrary, causes premature separation of sister chromatids. Wapl depletion does not affect phosphorylation of SA2 or the localisation patterns of other prophase pathway determinants like Plk1, Aurora B or condensin I, showing that Wapl does not act by modulating these proteins or their phosphorylation of cohesin. Importantly, Wapl controls cohesin association with chromosomes even in interphase cells and FRAP (fluorescence recovery after photo bleaching) studies showed Wapl depletion causes a change in the kinetics of this association in a way that the residence time of cohesin on chromatin was increased. These experiments established the role of Wapl as a major determinant of resolution of arm cohesion and showed that it acts independently of the phosphorylation/dephosphorylation events that also regulate the prophase pathway.

Although the mechanisms of the prophase pathway have been reasonably well-studied, the precise physiological role of the process is not yet clear. It has been suggested that the pathway might cause untangling of the chromosome arms and this could aid in the

condensation of chromosomes. Additionally, arm resolution might also shift the equilibrium of topoisomerases toward decatenation rather than catenation (Holm and Rosenstrom, 2010). However, convincing evidences in favour of either of these proposed roles of the prophase pathway is lacking. Another speculation is based on the role of cohesin in gene expression. The prophase pathway removes majority of the cohesin molecules from vertebrate chromosomes and as discussed earlier, the process involves dissociation of the complex without being cleaved or degraded. This may spare the cells from having to resynthesize the cohesin proteins before the subsequent G1 phase when the complex is needed for regulation of gene expression (Wendt et al., 2008).

1.6.3.2 *Stabilisation and removal of centromeric cohesin*

Centromeric cohesin in vertebrate cells and cohesin along entire chromosomes in yeast persist until biorientation is achieved and are removed only during metaphase-to-anaphase transition. Following this bipolar arrangement of sister chromatids, the spindle checkpoint is inactivated which triggers progression of the cells to anaphase. The protein Sgo1 (shugoshin) has been shown to be essential for protection of centromeric cohesin from removal by the prophase pathway (Salic et al., 2004) (Kitajima et al., 2005) (McGuinness et al., 2005). Sgo1 localises at centromeres as cells enter mitosis and persist until anaphase and depletion of the protein from mammalian cells causes premature cohesin dissociation and separation of sister chromatids without the cells proceeding to anaphase. Thus, in these cells centromeric cohesion is lost without the activation of APC/C and separase showing the protective role played by Sgo1.

Sgo1 also interacts with the PP2A complex at the centromeres (1994; Tang et al., 2006) (Murshudov et al., 1997) and the complex has been shown to dephosphorylate the SA2 subunit of cohesin *in vitro*. Thus, recruitment of PP2A at centromeres by Sgo1 is thought to protect cohesin from Plk1-mediated phosphorylation that leads to dissociation of cohesin from chromatin. Consistent with this, expression of a non-phosphorylatable mutant of SA2 in mammalian cells overcomes the defect in cohesion caused by Sgo1 depletion. A crystal structure of the N-terminal of human Sgo1

(residues 51-96) in complex PP2A holocomplex showed that the corresponding region of Sgo1 forms a single helix that dimerises to form a parallel coiled-coil (Xu et al., 2009). Each of the Sgo1 helices interacts with one PP2A complex through the regulatory and scaffolding subunits of the phosphatase. Importantly, the corresponding binding interface between the proteins was shown to be important for the protection of centromeric cohesin.

The centromeric localisation of Sgo1 is dependent on the conserved spindle checkpoint protein Bub1 (Tang et al., 2004) (Kitajima et al., 2005). Repression of Bub1 causes complete displacement of Sgo1 from centromeres and a loss sister chromatid cohesion. Loss of Bub1 further leads to a weaker localisation of Sgo1 along the chromosome arms. The conserved heterochromatin protein HP1 has also been implicated in recruiting Sgo1 to the centromere (Yamagishi et al., 2008). A direct interaction between the proteins has been observed and an Sgo1 mutant unable to interact with HP1 is not able to localise to the centromeres. Another protein important for centromeric cohesion is Aurora B (Ipl1 in budding yeast) and inhibition of the protein leads to loss of centromere cohesion. A recent study showed Aurora B phosphorylates the shugoshin family protein Sgo2, which can then bind to and recruit PP2A to the centromeres.

Cohesin associated to chromosomes in budding yeast and to the centromeres in vertebrate cells persist until metaphase-to-anaphase transition in order to facilitate biorientation. Prior to anaphase onset, inactivation of the spindle checkpoint causes the anaphase promoting complex/cyclosome (APC/C) to ubiquitinate the proteins securin and cyclin B subunit of Cdk1, triggering their degradation (Michaelis et al., 1997). An intact form of securin maintains the protease separase in an inactive form by binding to the protease during interphase and early mitosis (Hornig et al., 2002) (Waizenegger et al., 2002). However, targeting securin for degradation activates the separase, which then cleaves the Scc1 subunit of cohesin. Scc1 has two cleavage sites for separase both of which have been showed to be important. Cleavage of Scc1 is also facilitated by its phosphorylation by Plk1 (Alexandru et al., 2001) and leads to the separation of sister chromatids and their subsequent segregation during anaphase. Scc1 fragments generated after it is cleaved undergo ubiquitylation based degradation (Rao et al., 2001). In

budding yeast, separase activity has been shown to be sufficient for onset of anaphase and chromosome segregation in contrast to vertebrate cells, which also depend on the prophase pathway. Nonetheless, the separase pathway of cohesin removal is essential in the vertebrate cells for them to segregate chromosomes and absence of separase causes segregation defects (Hauf et al., 2001). However, unlike in the budding yeast, it is not yet clear whether separase activity by itself is sufficient for removal of cohesin from the centromeres in vertebrate cells.

1.7 Summary of Wapl/Wpl1 functioning and aims of the project

Budding yeast Wpl1 and its vertebrate orthologues are conserved eukaryotic proteins that share a number of common characteristics like a similar C-terminal WAPL domain, ability to associate with chromatin through cohesin and formation of a subcomplex with Pds5. Nevertheless, there exist differences between the Wapl orthologues of lower eukaryotes like budding yeast and those of vertebrates. In vertebrates, *wapl* depletion causes a delay in removal of cohesin from chromosome arms whereas overexpression of the protein leads to premature cohesion loss. Thus, the vertebrate orthologue clearly plays a major role in the prophase pathway of cohesin removal by somehow modulating cohesin-DNA interactions during early mitosis. In budding yeast, however, cohesin is removed from chromosomes in a single step during metaphase-to-anaphase transition mainly through the proteolytic cleavage by separase. So, the Wpl1 orthologue does not seem to have an obvious role in the process of cohesin removal and in addition, deletion of the corresponding gene in budding yeast leads to partial cohesion loss, which is contrary to the situation in vertebrates. Subsequently, studies in budding yeast identifying suppressors that bypass the lethality of *ECO1* mutants have yielded key insights into the functioning of the Wpl1 orthologue. The results showed that Eco1 is essential to acetylate the cohesin subunit Smc3 at two conserved lysines during S phase. Defects in this acetylation causes lethality that can be overcome upon deletion of the WPL1 gene. These observations have prompted the hypothesis that Wpl1 has a destabilising effect on cohesin that helps maintain the complex in a so-called non-cohesive or inactive form prior to duplication of the chromosomes. During S phase, Eco1 is activated in a replication dependent manner and acetylates cohesin at Smc3,

which is thought to convert cohesin to a cohesive state by opposing the Wpl1-mediated destabilisation.

The insights gained from these studies in budding yeast, thus, show Wpl1 might have a role similar to that of the vertebrate Wapl, that of promoting dissociation of cohesin from DNA, and thus a conserved mechanism of its functioning. However, the timing in the cell cycle when the function of the protein is more important varies among the corresponding orthologues, which might explain difference between the Wapl depletion and Wpl1 mutant phenotypes in the respective organisms. These studies have led us to a better understanding into the regulation of the process although some equally important and outstanding issues still remain unaddressed. One of the most important of these is what is the mechanism of destabilisation of cohesin and/or cohesin-DNA interactions by Wapl/Wpl1. Another key question is how acetylation of Smc3 helps in establishment of cohesion and how it overcomes Wpl1 mediated destabilisation (Figure 1-7). To answer these questions, a detailed characterisation of the protein Wpl1 has been carried out using structural and biochemical methods.

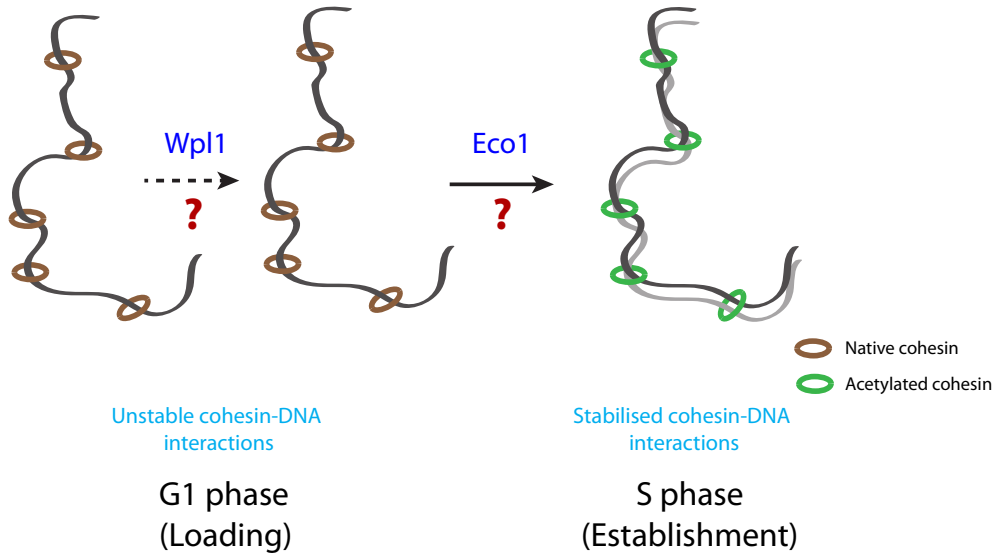


Figure 1-7: Regulation of cohesion by Wpl1 and Eco1

Chapter 2. Materials & Methods

2.1 Cloning

The *WPL1* genes of both *S. cerevisiae* and *A. gossypii* corresponding to full-length proteins and WAPL domains were PCR amplified from the respective genomic DNA of the organisms. The genes were cloned into a modified pET-28b(+) vector with a N-terminal His-tag sequence followed by a TEV (*t*obacco *e*tch *v*irus) protease cleavage site. The DNA sequence corresponding to the Smc3-ATPase (residues1-190 and 1047-1231 linked by the sequence SGGSGG) PCR amplified from the *S. cerevisiae* genome and was cloned into a modified pET-28b(+) vector with a N-terminal GST tag sequence.

Amplification of the genes was carried out using Velocity DNA polymerase (Bioline) according to the following cycling conditions:

Steps	Description	Temperature	Time	Cycles
1	Initial denaturation	98 °C	2 minutes	
2	Denaturation	98 °C	30 seconds	
	Annealing	55 °C	30 seconds	25 cycles
	Extension	72 °C	30 seconds/kb	
3	Final extension	72 °C	10 minutes	

The amplified products were checked using an agarose gel and were purified using Qiagen PCR purification kit. The purified DNA fragments were then subjected to restriction digest for a period of 6 hours to overnight.

The digested DNA was ligated to a suitably digested and PCR purified vector using the T4 DNA ligase (NEB) and appropriate buffer in a final reaction volume of 5 to 10 μL. The reactions were incubated overnight at room temperature. 0.5-1.0 μL of the ligation mix was then transformed into an aliquot of XL-1 electro competent cells. The cells were incubated at 37 °C for 1 hour and plated on LB agar containing the suitable antibiotic.

Following an overnight incubation at 37 °C, the colonies obtained were screened by colony PCR using MangoTaq polymerase (Bioline). 5 mL overnight cultures of the same colonies were simultaneously set up from which the respective plasmids were isolated to perform sequencing. Sequencing PCR was carried out with the help of BigDye mix using either T7 forward and reverse primers or gene specific primers. The reactions were cleaned up using ethanol/EDTA precipitation method. The samples were then submitted to the LRI sequencing facility to carry out sequencing.

2.2 Mutagenesis

A PCR-based site-directed mutagenesis approach was used to introduce desired mutations in the gene of interest. Primers with suitable mutations were designed using the Stratagene Primer Design software. PCR amplifications were performed using the *Pfu* Turbo high fidelity polymerase and company supplied buffer.

The PCR reactions were set up as follows:

10X Buffer	5 μL
Plasmid (containing target gene)	20 ng
Forward primer	10 pmoles
Reverse primer	10 pmoles
dNTP mix	1.25 μL of 10 mM each
H_2O	up to 50 μL
Pfu polymerase	2.5 units

The following steps were used for the amplification:

Steps	**Description**	**Temperature**	**Time**	**Cycles**
1	Initial denaturation	95 °C	1 minutes	1 cycle
2	Denaturation	95 °C	30 seconds	
	Annealing	55 °C	1 minute	18 cycles
	Extension	68 °C	2 minutes/kb	
3	Final extension	72 °C	10 minutes	1 cycle

After amplification, 10 units of *Dpn*I enzyme was added to the contents of the tube and mixed by pipetting. The tubes were incubated for 1 hour at 37 °C followed by transformation of 2 μL of the digested product into XL1 cells and plated on appropriate antibiotic plates. Cells from a few colonies (5-10) were grown overnight in 5 mL culture volumes, the DNA from the cells extracted and sequenced to verify the mutations.

2.3 Expression and purification of the proteins

The proteins were overexpressed in an *E. coli* BL21 (DE3) RIL expression strain. Cultures of cells expressing the Wpl1 constructs were grown until an A_{600} of 0.6 units at 37°C and induced with 0.5 mM IPTG at 18°C for 16-18 hours. Cells were lysed by sonication for 5 x 30s (for 2 L culture) and centrifuged at 34,000xg for 45 minutes in a buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 10 mM ß-ME (buffer 1) supplemented with protease inhibitor tablets (Roche). The supernatant was incubated with Ni-Sepharose 6 Fast Flow beads (GE Healthcare) pre-washed in buffer 1 for 1 hour at 4°C before washing with 10 column volumes of buffer 1. Elution was carried out using buffer 2 (50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole, 10 mM ß-ME). The His-tag was then cleaved by overnight incubation at 4 °C with TEV protease at a ratio of 1:50 of protease to protein. Size exclusion chromatography using a Superdex 200 column (GE Healthcare) was then performed as a final purification step in the buffer 40 mM HEPES 7.5, 200 mM NaCl and 1 mM DTT. For selenomethionine (SeMet) labeling of $AgWpl1^{184-561}$, the recombinant plasmid was transformed into a B834+ strain of *E. coli* and grown in minimal LeMaster medium containing L-selenomethionine. The SeMet-labeled protein was purified as for the native protein.

The ScSmc3-ATPase domain, which was expressed as a GST-fusion protein the cultures were grown until an A_{600} of 0.4 units at 37 °C and induced with 0.5 mM IPTG at 18°C for 16-18 hours. Cells were lysed in a buffer containing 50 mM Tris pH 8.0, 100 mM NaCl, 0.1% Tween-20, 1 mM DTT (buffer A) and centrifuged at 34,000 x g for 45 minutes. The cleared lysate was incubated with glutathione Sepharose beads (GE Healthcare) pre-equilibrated in buffer A, for 2 hours at 4 °C. The beads were then

washed with 10 column volumes of buffer B (50 mM Tris pH8.0, 100 mM NaCl, 1mM DTT) and the protein was finally eluted in the buffer C (50 mM Tris pH 8.0, 100 mM NaCl, 40 mM reduced glutathione, 1 mM DTT). Size exclusion chromatography was then carried out using a using a Superdex 200 column in a buffer containing 40 mM HEPES pH 8.0, 150 mM NaCl, 1mM DTT.

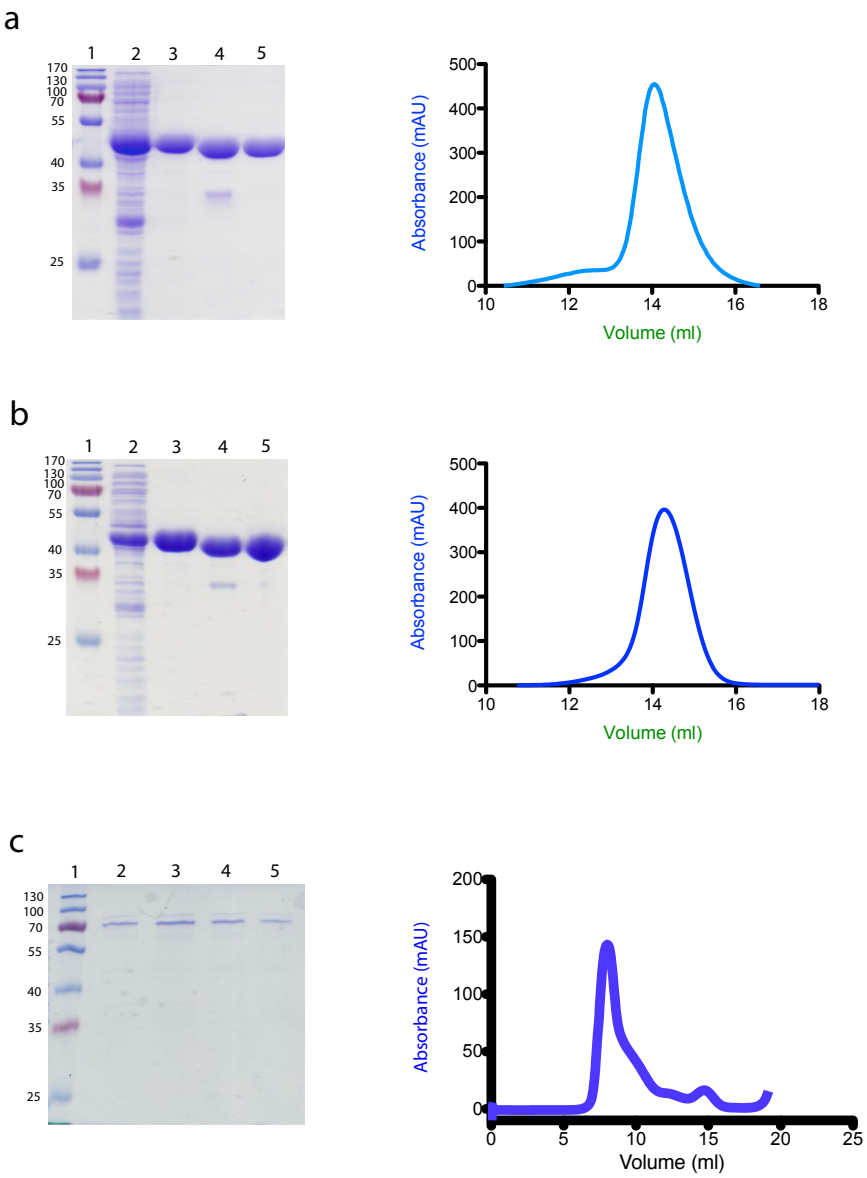


Figure 2-1: Gels and traces

ScWpl1$^{259-647}$ (a) and AgWpl1$^{184-561}$ (b) purification profiles and gel filtration traces (right side) respectively. Both the gels have been loaded in the order – marker (lane 1), lysate (lane 2), Ni-NTA purified (lane 3), TEV digested (lane 4), and gel filtration purified (lane 5). (c) Gel showing Smc3-ATPase protein fractions obtained after gel filtration (trace on the right). All the gel filtrations shown have been performed using 24 mL Superdex 200 column.

2.4 Fluorescence polarisation assay

2.4.1 Principle

Polarised light selectively excites fluorophore molecules whose absorption transition vectors are aligned parallel to the electric field of the light. The excited fluorophores, however, get randomly oriented due to their mobility or tumbling in solution, resulting in the emission of relatively depolarised light. When a fluorophore is bound to a larger molecule, the mobility decreases because of the added mass and light emitted from this bound form is relatively more polarised than that emitted from the free fluorophore molecules. Thus, measuring the polarisation of the emitted light provides a direct means to analyse binding between two molecules, one of which (preferably the smaller one) is attached to a fluorophore.

In order to measure the extent of polarisation, intensities of light emitted in planes both parallel ($F_{||}$) and perpendicular ($F_{\perp}$) to the direction of polarisation is measured. The polarisation (P) is measured according to the following equation:

$$P = \frac{I_{II} - I_{\perp}}{I_{II} + I_{\perp}} \tag{2.1}$$

The extent of polarisation of emitted light can also be measured in terms of anisotropy (r). Anisotropy takes into account the intensity of emitted light along the second possible perpendicular plane relative to the plane of polarisation and therefore, is considered a more accurate measurement of the extent of polarisation. For this reason, to calculate anisotropy, the perpendicular term in the denominator of Equation 2.1 is multiplied by two. Hence, the formula for calculation of anisotropy would be:.

$$r = \frac{I_{II} - I_{\perp}}{I_{II} + 2I_{\perp}} \tag{2.2}$$

Two factors important in determining the fluorescence anisotropy are the rotational correlation time (ϕ) and the fluorescence lifetime (τ). The rotational correlation time

depends on the solvent viscosity and volume of the molecule according to Stokes equation:

$$\phi = \frac{\eta V}{RT} \tag{2.3}$$

The relation between anisotropy (r), ϕ and τ is given by the following equation:

$$r = \frac{r_0}{1 + \left(\frac{\tau}{\varphi}\right)} \tag{2.4}$$

where r_0 is called the intrinsic or limiting anisotropy.

The fluorescence lifetime represents the time for which a fluorophore remains in the excited state following incidence of the polarised light and is a characteristic property of the nature of the fluorophore (τ for fluorescein is 4 ns). Thus, the rotational correlation time of a complex (e.g., between fluorophore-bound ligand and the interacting protein) and fluorescence lifetime of the fluorophore needs to be considered while designing a fluorescence polarisation experiment.

2.4.2 Protocol used

The binding of fluorescein-labelled Smc3 peptides to purified AgWpl1$^{184\text{-}561}$ was checked using the FP assay. The protein was dialysed into a buffer containing 20mM Tris pH 7.5, 20 mM NaCl, 0.5 mM TCEP and the same buffer was used to resuspend and dilute the peptide stocks. The peptide concentration used for the assay was 50nM while varying the protein concentration from 0 to 300μM; the total volume of the assay was 20μL. The anisotropy readings were recorded after 1 hour incubation of the protein and peptides at 4°C to allow attainment of equilibrium. Readings were taken using the Saffire^2 (Tecan) microplate reader in Fluorescence polarization mode. Binding experiments were carried out with different batches of protein and anisotropy readings from three independent (triplicates) experiments for each concentration of the protein of

a single batch were used for anisotropy calculations. The formula used for the calculations was

$$Y = A_f + (A_b - A_f)\left[\frac{X}{K_D + X}\right] \quad (2.5)$$

where, X is the protein concentration, Y is the observed anisotropy, A_b is anisotropy at saturation due to the bound ligand, A_f is anisotropy from the free ligand and K_D is the equilibrium dissociation constant in same units as X. The data was analysed and the graphs were plotted using the software GraphPad (Prism).

2.5 Biolayer interferometry assay

2.5.1 Principle

Biolayer interferometry (BLI) is a technique based on optical interference to analyse macromolecular interactions. In this method, one of the binding partners is immobilised on the biosensor tips and the other macromolecule is maintained in solution. The assay uses special biosensors having an optical layer at their tips used as an internal reference to compare interference patterns. White light is allowed to pass through the biosensor, which gets reflected from two different surfaces – one being the interface between the biosensor and the optical layer while the other surface is the edge of the biosensor tip in contact with a solution containing one of the proteins (Surfaces 1 and 2; Fig. 2-1). The same wavelengths of the white light reflected from the two surfaces are analysed separately, which are then combined to obtain an interference (or interferometry) pattern. For example, interference of yellow light from both the surfaces (yellow channel) is analysed and similar analyses is simultaneously done for other wavelengths as well. These individual interferences of different wavelengths among themselves are then used to construct an interference pattern. Now, if the thickness of the surface at the biosensor tip increases (due to binding of the macromolecule in solution to the immobilised ligand on the biosensor tip), the path length for the light reflected from this surface increases compared to that reflected from the optical interface (the reference; which remains the same) (Fig. 2-1). Change in the path length changes the interference

among the rays of the same channel and thus changing the overall interferometry pattern. As more molecules attach to the surface, the thickness increases resulting in increased shift in the interferometry pattern. Conversely, dissociation of molecules from the surface results in decrease in the shift. Thus, the association and dissociation of molecules at the surface can be monitored in real-time and the kinetics of binding reactions can be calculated.

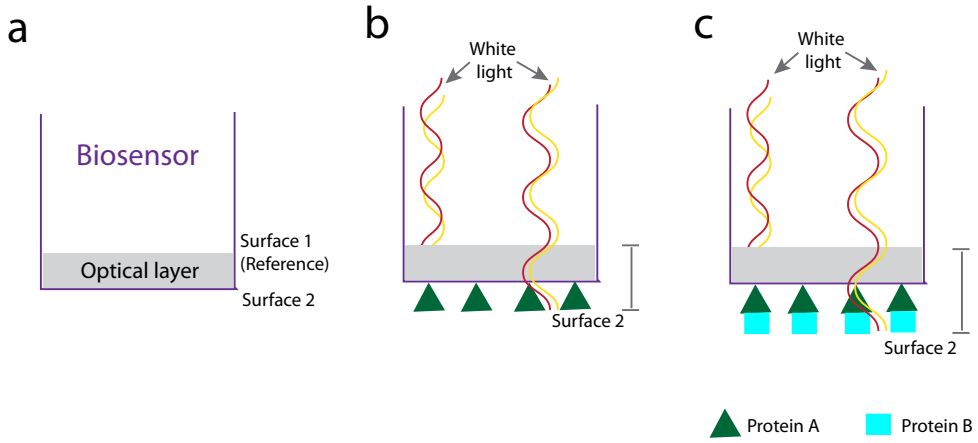


Figure 2-2: Principle of biolayer interferometry

A biosensor tip showing the optical layer and the two reflecting surfaces (a). An increase in the surface thickness due to binding of a molecule to a ligand immobilised on the surface of the biosensor increases the pathlength of the reflected light from Surface 2, causing a shift in the interferometry pattern (b & c).

2.5.2 Protocol used

Binding studies between the proteins ScWpl1$^{259-647}$ and GST-tagged ScSmc3-ATPase were performed using the biolayer interferometry experiments with the help of an OctetRed (ForteBio) instrument. Proteins used in this assay were dialysed into a buffer containing 40 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT. Purified ScWpl1$^{259-647}$ was amine-coupled to the biosensor tips at a concentration of 12.5 µg/mL. Best results were obtained when coupling was carried out at pH 5.0. The protein ScSmc3-ATPase was maintained in solution at concentrations of 0, 25, 50, 100, 200, 400, 800 and 1200 nM. The biosensor tips containing the immobilised ScWpl1$^{259-647}$ protein were then dipped into the wells containing different concentrations of the Smc3-ATPase

simultaneously to measure the kinetics of the binding reaction between the proteins. The same protocol was repeated for wild type as well as mutant proteins that were used. The kinetics were measured using provided software.

2.6 Peptide arrays

The ATPase head domain of Smc3 (residues 1-190 & 1045 -1231) from A. gossypii was arrayed on a cellulose membrane as peptide spots, each comprising of 21 amino acids with a shift of 3 amino acids between successive peptides. The membrane was pre-incubated in a buffer containing 40 mM HEPES 7.5, 150 mM NaCl, 0.1% Tween-20, 0.25 mM TCEP (incubation buffer) and then 50 nM AgWpl1184-561 added, and incubated overnight at 4 °C. Blocking carried out in 40 mM HEPES 7.5, 150 mM NaCl, 0.1% Tween-20, 0.5% non-fat milk, 0.25 mM TCEP (blocking buffer) for 1 hour at room temperature (RT). The array was then washed with incubation buffer for 3 X 10 minutes followed by probing with anti-AgWpl1184-561 diluted 1:10,000 for 1 hour at room temperature. The membrane was washed again with the incubation buffer for 3 X 10 minutes before adding the secondary antibody (goat anti-rabbit HRP conjugated, DAKO; 1:2000 dilution). The array was visualised using the ECL detection kit (GE Healthcare).

2.7 *In vivo* assays

The steps followed for the in vivo assays were as follows:

I. Cloning of the *S. cerevisiae WPL1* gene along with its own promoter into the plasmid pYIPLac204 followed by epitope tagging of the construct with HA-tag.

II. Deletion of the endogenous copy of *WPL1* gene from *eco1-1* mutant yeast cells (refer to results and introduction for the rationale)

III. Transformation of the pYIPLac204 carrying the *WPL1* gene with its promoter into *eco1-1*/*WPL1Δ* cells

Step I: Cloning of *WPL1* along with its promoter

The *WPL1* gene along with its promoter, which consists of a 200 base pairs sequence upstream of the gene, were amplified using forward and reverse primers having SalI and BamHI restriction sites incorporated, respectively. The reverse primer, in addition, also contained a NotI restriction site (Fig. 2.2). The sequence was amplified using PCR steps as described earlier. An HA tag sequence digested with NotI restriction enzyme was then inserted into the C-terminal NotI site of the WPL1 construct to facilitate its detection.

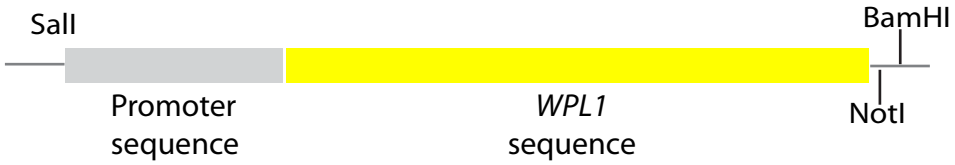


Figure 2-3: The *WPL1 in vivo* construct

Step II: Deletion of the *WPL1* gene from *eco1-1* strain

The marker gene *LEU* was used to disrupt the endogenous copy of *WPL1* gene from the eco1-1 mutant strain. The *LEU* gene (inserted into a pBS plasmid) was amplified using primers, which had sequences corresponding to the *WPL1* gene flanking *LEU* sequence.

The following primers were used for amplification of *LEU* with *WPL1* flanking sequences.

Forward primer:
TCGCAAAACGAAACCATCTTCTTACCCTAAAGCATCCTGTTTCTGAAAAAtaat acgactcactataggg

Reverse primer:
GTTGCCAGCAGGGTGAAGATGAAGCCAGGCTATGTTCAATGTATGCTTTCTat taaccctcactaaaggg

The bases in capital and small letters in the above primers show sequence complementary to *WPL1* and *LEU* respectively.

Forward primer for checking deletion:
GTTGAGGAGGCTTTCTGAGC

Reverse primer for checking deletion:
CGCTAGAAGGCTCATCAAG

Step III. Transformation

The vector containing the *LEU* construct was transformed into cells from *eco1-1* strain. The genomic copy of the *WPL1* gene was replaced by the *LEU* by homologous recombination. The pYIPLac204 plasmid carrying a HA-tagged *WPL1* gene was then transformed into these cells. The protocol for yeast transformation used is given below.

Buffers used:

(a) 1X TEL:
10 mM Tris-HCl pH 7.5
0.1 mM EDTA
100 mM Lithium acetate

(b) 10X TEL:
100 mM Tris-HCl pH 7.5
1 mM EDTA
1M Lithium acetate

(c) 50% PEG 3350 (in sterile water)

(d) TELP (per mL):
100 µL 10X TEL
100 µL water
800 µL 50% PEG3350

Protocol used:

A 100 mL culture of the *eco1-1* strain was grown until mid logphase and centrifuged at 3000 rpm for 5 minutes. The cells were washed with 1 mL deionised water and spun down in again. The cells were then washed with 1 mL TEL, spun down and then resuspended in 100 µL TEL. 1 µg of amplified *LEU* was mixed with 2 µl of a 10 mg/mL single stranded salmon sperm carrier DNA and 300 µl TELP. 50 µL of the cell suspension was added to this mix followed by a short vortex (10 seconds). The cells were then incubated at 25 °C for 4 hours followed by a heat shock at 42 °C for 15 minutes. The cells were then spun down at 6,000 rpm for 2 minutes, washed in 1 ml of 1 M sorbitol and plated on selective media (here LEU drop-out media). The plates were incubated at 25 °C and colonies appeared after 3 days. Transformants were checked for the correct integration of *LEU* by PCR.

For transformation of pYIPLac204 plasmid carrying *WPL1* genes (with HA-tag sequence; either wild-type or mutants), the plasmid was linearised in the TRP1 marker using EcoRV enzyme. The linearised plasmid was then transformed into *eco1-1*/WPL1Δ strain using the same procedure described before. The cells were then plated on TRP deficient medium. The expression of the tagged Wpl1 protein was verified by Western blot using an anti-HA antibody.

Yeast DNA preparation for checking deletion of *WPL1* by PCR genotyping:

Buffers:

(a) SCE:
1 M Sorbitol
0.1 M Sodium citrate pH 7.0
60 mM EDTA

(b) SCE/ME/Zymolase:
8 μL/mL βME and 2 mg/mL Zymolase T-20 added to SCE

(c) SDS solution:
100 mM Tris-HCl pH 9.0
50 mM EDTA
2% SDS
(d) 5 M Potassium acetate

Freshly patched cells were taken with a tooth-pick and resuspended in 200 μL SCE/ME/Zymolase. The cells were incubated at 37 °C shaker for 45 minutes. 200 μL of the SDS solution was then added and incubated at 65 °C for 5 minutes. 200 μL of 5 M potassium acetate was added to the mix and centrifuged for 10 minutes at 14,000 rpm. 350 μL of the supernatant was taken and to this 800 μL of ethanol was added at room temperature. The mixture was spun for 2 minutes at 6000 rpm, the ethanol was pipetted

out and the pellet rinsed with 70% ethanol followed by air-drying. The DNA was finally dissolved in 200 μL water and 0.5 μL of this was used for PCR. A standard colony PCR amplification protocol was used. Presence of the LEU gene can be checked by the presence of a 1.5 kb fragment in the amplified product.

Western blot to verify WPl1 expression:

Proteins for Western analysis were extracted from the yeast cells using the NaOH lysis method. Cells from 5 mL of overnight culture were harvested by centrifugation at 3000 rpm for 5 min at 4 °C. The cells were resuspended in 1 mL of ice-cold deionised water and centrifuged. The cells were again resuspended in 150 μL of NaOH solution (prepared by mixing 1 mL of 2M NaOH with 80 μL of β-mercaptoethanol) and spun down at 13000 rpm for 1 minute. The pellet was dissolved in 2X SDS loading buffer and boiled at 95 °C for 5 minutes. 5-10 μL of the sample was loaded onto a SDS-PAGE gel.

The resolved proteins were blotted onto a nitrocellulose membrane pre-soaked in a transfer buffer (14.4 g/L Glycine, 3 g/L Tris-base, 0.02% SDS, 10% v/v methanol). A wet transfer was performed at 100V for 2 hours at 4 °C. Following the transfer, the membrane was washed with PBS and blocked for 20 minutes using the blocking buffer (5% milk and 0.1% Tween 20 in PBS). The membrane was then incubated with the primary antibody (α-HA), diluted up to 1:5000 in blocking buffer, for 45 minutes at room temperature. Following the incubation, the membrane was washed 4 x 5 minutes in PBST (PBS and 0.1% Tween 20) and then incubated in the HRP-conjugated secondary antibody that was diluted 1:5000 in the blocking buffer. The blots were washed again for 4 x 5 minutes with PBST and developed using ECL mixture (1:1 ratio of solutions 1 and 2; Amersham).

Chapter 3. Theory of protein crystallography

X-ray crystallography has been the key structure determination tool and remains the method of choice for gaining detailed insights into the three-dimensional structure of proteins. The underlying principles and theories of crystallography have been briefly described below with an emphasis on the anomalous diffraction phasing technique that was used for solving the WAPL domain crystal structure.

Macromolecules are made up of large number of covalently bonded atoms with interatomic distances in the range of a few angstrom units (Å). Obtaining atomic resolution images of these macromolecules makes it necessary to use X-rays with wavelengths optimally within the range 0.5-1.6 Å. Availability of an X-ray lens and thus an X-ray microscope, on the lines of the normal light microscopes, would have simplified the whole process of visualising atomic structures. However, it is improbable to design lenses suitable to converge radiations of such small wavelength in order to generate an image. In addition, when X-rays diffract, the phase information is lost and needs to be obtained indirectly followed by mathematical image reconstruction.

Obtaining good-quality crystals of a target protein is a prerequisite for any X-ray diffraction experiment aimed to decipher an atomic resolution structure. Crystals are formed of numerous individual copies of a protein molecule arranged in an ordered three-dimensional array, which help not only to amplify the diffraction signal but also to minimise information loss due to radiation damage. Presence of large number of copies serves to overcome the signal loss arising from damage to a small proportion of the individual molecules. However, crystallising a protein may often prove to be a limiting step due to unpredictable outcomes of a crystallisation experiment. In addition, the fact that only limited amounts of the purified proteins can be obtained in many cases further aggravates the problem. Nonetheless, crystallising a protein is an absolute necessity to initiate a structure solution process and often involves a hit and trial method to identify suitable conditions for crystal growth.

3.1 Crystallisation

In order to grow crystals, a purified protein (> 90% pure ideally) needs to be maintained at a supersaturated state. A sufficiently high level of supersaturation is necessary to overcome an energy barrier that allows formation of the crystal nuclei. An excessively high protein concentration, however, is detrimental to crystal growth and instead pushes the equilibrium toward random association of the protein molecules giving rise to large disordered aggregates rather than ordered crystals. On the other hand, an adequate level of protein supersaturation leads to slow and ordered nuclei formation. Under favourable conditions, formation of these nuclei would result in a decrease in the protein concentration of the surrounding solution, bringing it to a metastable state, ideal for an ordered interaction among protein molecules or small aggregates giving rise to crystals (Fig. 2-1). Supersaturation of the protein is achieved by mixing the protein (in a suitable buffer) with a crystallisation solution normally containing a precipitant in addition to a salt and a buffer component. The precipitant shields the water molecules thereby increasing the effective concentration of the protein in the drop and also aids in slow diffusion of water from the drop (due to its relatively lower concentration in the drop initially).

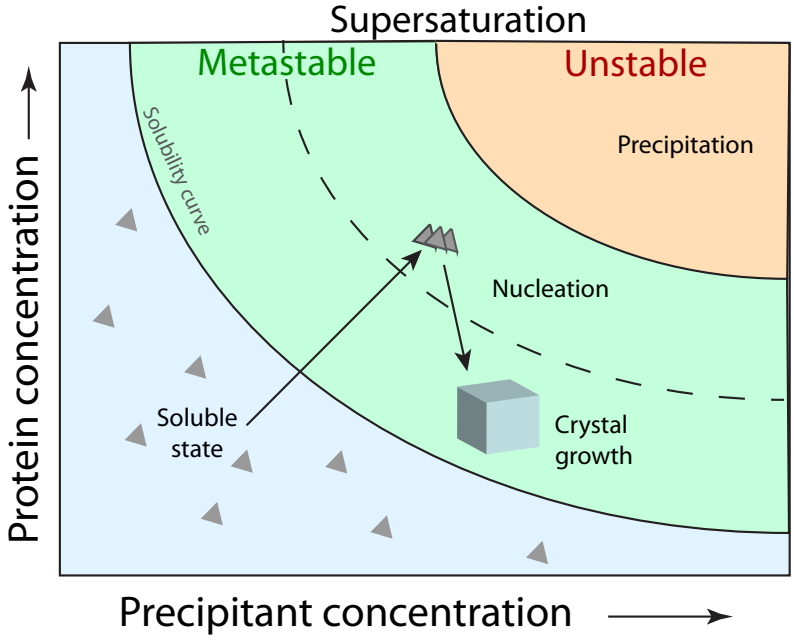


Figure 3-1: Behaviour of a protein in solution

The success of a crystallisation experiment also depends on the physico-chemical properties of a protein. Proteins are large and flexible molecules with irregular surfaces, factors that limit the number of interactions possible between neighbouring molecules. Furthermore, protein surfaces exhibit localised surface charge patterns, which also vary between different proteins making them behave uniquely in a crystallisation experiment. Surface properties of proteins, thus, decide how protein molecules interact, a factor crucial in determining the success of the crystallisation screens or trials.

One of the most widely used methods of protein crystallisation involves the vapour-diffusion technique and is used either in a hanging drop or a sitting drop format. In both the set-ups, a protein solution at a suitable concentration is mixed with an equivalent amount of the precipitant solution. In the hanging-drop method, few microliters (1-5 μL) of the protein and precipitant solutions are placed on a siliconised glass cover-slip, which is then inverted on a well containing a relatively large amount (0.5-1.0 mL) of the precipitant solution. In the sitting drop format, as the name suggests, the drops are set onto a depression of a small elevated platform placed within the well. In both the cases, the wells along with the drops containing the protein are sealed properly to avoid evaporation and to allow vapour diffusion to proceed without any external influence. The well-solution contains a higher concentration of the precipitant (almost double in most cases) compared to the protein drop and causes water-vapour from the drop to diffuse out into the well causing a gradual increase in the protein concentration until an equilibrium is attained.

3.2 Crystal packing

Crystals of an organic material like a protein consist of molecules arranged in a three-dimensional periodic array. When protein molecules precipitate from their supersaturated solutions, they tend to reach the lowest free energy state, achieved by forming a regular arrangement of the molecules that ultimately gives rise to crystals. Crystals in general are characterised by distinct edges and angled planes forming the boundaries. Differences in the shapes of the planes and angles between them give rise to different crystal forms and this outer morphology reflects the way its contents are

arranged. The diffraction pattern obtained from a particular crystal depends on the nature of its contents and their packing in space. Therefore, an understanding of the general periodic patterns, or in other words, the symmetry of the crystal, is of primary importance to ultimately be able to elucidate its atomic structure.

3.2.1 Crystal lattice

The basic repeating unit of a crystal is known as a 'unit cell', an imaginary block whose edges are denoted by vectors **a**, **b** and **c** (in the directions x, y, and z respectively) with angles α, β, and γ between them. The planes formed by the unit cell vectors **a** and **b** is depicted as C, by **b** and **c** as A and by **c** and **a** as B (Fig. 2-2a). The imaginary boundaries of the unit cells form repeating grids, the assembly of which give rise to the entire crystal lattice.

Unit lattices may be differentiated into primitive or centred and in three-dimensional space, there can be a total of six primitive and eight centred lattices that together make up the 14 Bravais lattices. A primitive lattice contains one lattice point per cell (by combining 8 fractional lattice points at each of the vertices) and the unit lattices within such a crystal are related by lattice translations, which are whole number multiples of its unit cell vectors. A centred lattice, on the other hand, contains more than one lattice points per cell. The additional points arise due to centering in this type of lattice, which may be located at one or more faces of the planes (face-centering, *F*) or the centre (body-centering, *I* or rhombohedral centering, *R*) of the lattice. Due to the centering, fractional translations may relate centred unit lattices within such a crystal (Figure 3-2b and c).

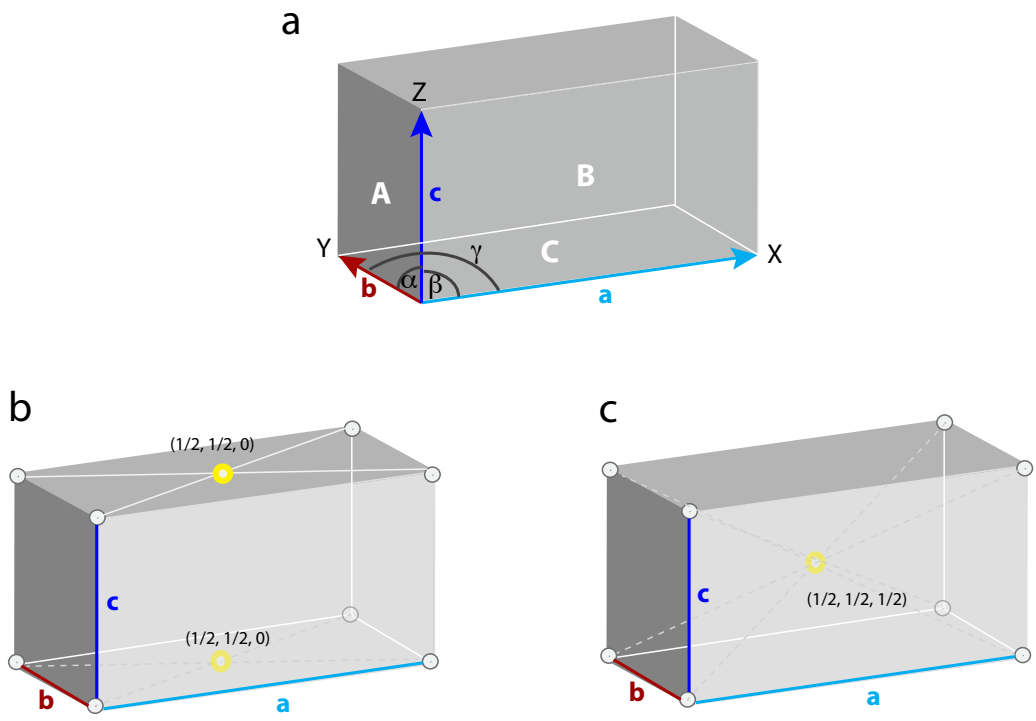


Figure 3-2: Unit lattice and centering

(a) A unit lattice showing the vectors **a**, **b** and **c** and the respective angles between them α, β and γ; also shown are the planes formed by these vectors, denoted as A, B and C. (b) A face centred (denoted as *F*; centering on the C plane) and (c) a body centred (denoted as *I*) unit lattice. The corresponding lattice translations in both the cases along the vectors **a**, **b** and **c** have been indicated in brackets.

3.2.2 Miller indices and the reciprocal lattice

Lattice planes in a crystal are defined by a set of indices denoted by h, k and l [represented as (hkl)] and are called the Miller indices. These are imaginary sets of planes, which cut through the **a**, **b**, and **c** axes of all the unit cells at fixed fractions and a set of planes defined by the same indices are parallel to each other. The fractions of the unit cell edges are expressed as whole numbers and represent the numerical values of h, k and l. For example, a set of planes with indices (**234**) divide the unit cell edge **a** into two, **b** into three and **c** into four parts. Many different sets of these planes with different indices exist which together define the entire crystal lattice.

The concept of reciprocal lattice is important to determine the direction of the diffracted beam and therefore, the positions of the diffraction spots or reflections for a crystal at a given orientation. The planes of the reciprocal lattice intersect an imaginary sphere, known as the Ewald sphere, of radius $1/\lambda$ drawn with the crystal at its centre. The points of intersection between this imaginary sphere and the reciprocal lattice represent the points where diffraction spots can be observed. At a specific orientation of the crystal, only a limited number of intersections are possible and thus, a limited number of diffraction spots. The diffractions in other directions can be observed by rotating the crystal as this will result in rotating the reciprocal lattice as well and the diffraction spots in other directions can be observed. The reciprocal lattice unit cell edges are generally denoted by the vectors **a***, **b*** and **c*** and have magnitudes that are reciprocal to that of the unit cell of the crystal.

3.2.3 Symmetry

Symmetry, in two-dimensions, may be defined as a single or a group of operation(s), which bring a molecule back to the same state as the original. However, in three dimensions, this may not always be true (e.g. in case of screw axis), so, in this case the symmetry operation should at least generate a translationally equivalent molecule. Copies of a molecule within a unit cell may be symmetrically related and the same can be applied to the contents of all the remaining unit cells making up the crystal. The basic symmetry operators that can exist are axes of rotation (with or without a translation), mirror planes, inversion centres and rotation inversion axes.

One of the most important prerequisites for a symmetry operation within a crystal is that it should not cause any change in the relative orientation of a molecule in the crystal. This requirement imposes a restriction on the nature of symmetry operation that the chiral protein molecules may undergo. For example, an inversion or mirror symmetry, which may well exist among non-chiral symmetry mates, are forbidden in case of proteins, as these operations would lead to a change of handedness of a protein. In addition, the requirement that the entire crystal lattice needs to be filled completely by the same repeating unit imposes further restrictions on the types of the rotational

symmetries. Accordingly, only 2-, 3-, 4- and 6-fold rotational symmetries can exist in crystals, as space cannot be fully occupied by repeating units formed by a 5- and 7-fold or higher rotational axes.

If a rotation is followed by a translation along the same axis, the operation gives rise to a screw axis. In a screw operation, a N-fold rotation is followed by a translation of *s*/N fraction of a unit cell edge, along the rotation axis. The values of *s* range from 1 to (N-1) and thus a 2-fold screw operation along Y-axis will involve a 180° rotation around Y followed by a translation of ½ of the magnitude of the vector **b** along the same axis. Similarly, for a 3-fold screw operation the translation component can have values 1/3 or 2/3 of the unit cell vector along the rotation axis.

Molecules within a protein crystal are often related by screw axis but the translational component of such an operation is not evident from the external morphology of a crystal. Thus at a macroscopic level, one is left only with rotation operations, combinations of which give rise to 32 point groups. These were derived from observing the external faces of a crystal and since normals of these external planes invariably intersect at a point, they were named as point groups. The 32 point groups can be categorised into only 7 crystal systems based on the symmetry operations and the restrictions imposed by the requirement of filling space. These are listed in Table 2-1. Crystals involving asymmetric or chiral molecules like proteins can have only 11 point groups.

Table 3-1: The crystal systems

Crystal system	Lattice conditions	Point group symmetry (unique axis)
Triclinic	$a\neq b\neq c$; $\alpha\neq\beta\neq\gamma\neq 90$	1
Monoclinic	$a\neq b\neq c$; $\alpha=\gamma=90$	2 (single 2-fold along b)
Orthorhombic	$a\neq b\neq c$; $\alpha=\beta=\gamma=90$	222 (three 2-folds along a, b and c)
Tetragonal	$a=b$; $\alpha=\beta=\gamma=90$	4 (one 4-fold along c)
Trigonal	$a=b$; $\alpha=\beta=90$; $\gamma=120$	3 (3-fold along c)
Hexagonal	$a=b$; $\alpha=\beta=90$; $\gamma=120$	6 (6-fold along c)
Cubic	$a=b=c$; $\alpha=\beta=\gamma=90$	23 or 432 (four 3-folds along space diagonals)

3.2.4 Asymmetric unit and non-crystallographic symmetry

The smallest repeating unit that can generate the entire unit cell contents by applying the symmetry operators is known as an asymmetric unit. For example, a triclinic unit cell, which has no symmetry, is made of only a single asymmetric unit whereas a monoclinic unit cell that has two identical symmetry-related molecules, consists of two asymmetric units. Thus, asymmetric units within a unit cell are related to each other by crystallographic symmetry.

The asymmetric unit may contain molecules not related by crystallographic symmetry to each other but are formed as a result of the protein crystallising as an oligomer. Such molecules are said to be related by local or non-crystallographic symmetry (NCS) and may be formed of any number of monomers (unlike symmetry mates that may have only 2, 3, 4, and 6 rotational symmetry related molecules). In contrast to crystallographic symmetry, the NCS-related molecules are often not the exact replica or symmetry equivalents of each other and may in some cases undergo interactions among themselves through some secondary structure motif. Interactions within NCS monomers often correlate with the oligomerisation state of the protein in solution.

3.2.5 Space group

The combinations of different symmetry operations together with the 14 Bravais translations give rise to 230 possible ways in three-dimensional space to pack crystals of symmetrical molecules. However, in case of chiral protein molecules, this number is reduced to 65 due to absence of operators like mirror planes or inversion centres. Space group is an efficient way to depict the combination of all the symmetry operations that exist within a crystal. A space group can be generally depicted in the form:

$$G \equiv T_B W_1 W_2 W_3$$

This is known as the Hermann-Mauguin space group notation and is a general way to depict the space group of a crystal. In this notation, T_B stands for the Bravais lattice

while W for the generating operators and denotes the symmetry operation along each of the three coordinate axes. For example, in the space group P222, P (for primitive) stands for the lattice type, 2 stands for the two-fold symmetry along all the three directions – the *x*-, *y*- and *z*-axes.

3.3 Properties of waves

The structure of a molecule and its three-dimensional arrangement within a crystal determine the intensities of the diffracted waves. In other words, a relationship exists between the nature and periodic distribution of the electron density in a crystal and its X-ray diffraction pattern. X-rays, which are electromagnetic waves, exert a force on the electrons when they hit the atoms and cause the electrons to oscillate with the same frequency as the incident radiation. The oscillating electrons emit radiation of the same frequency but with a different phase to that of the incident wave. The scattering of X-rays by a crystal is thus, a result of the addition of all the individual waves scattered from its numerous electrons. A basic understanding of the properties of waves is therefore important and the key concepts are discussed below.

Electromagnetic waves like that of light or X-rays are made of an electric and a magnetic field component, varying in perpendicular directions to each other. X-rays interact with charged particles like electrons only through its electric component and thus the magnetic field is not considered. The electric field (E) varies as a cosine function depending on the amplitude (A) and frequency (ν) of the wave and at any time '*t*' can be represented as:

$$E = A\cos 2\pi\nu t \tag{3.1}$$

Thus, the wave can also be expressed with respect to the distance of propagation (*x*) and wavelength (λ) as:

$$E = A\cos 2\pi\frac{x}{\lambda} \tag{3.2}$$

In equation 3.1, $2\pi\nu$ can be substituted with the angular velocity 'ω'. In addition, if we also include the phase of the wave, say ϕ, then the equation becomes:

$$E = A\cos(\omega t + \phi) \tag{3.3}$$

3.3.1 Addition of waves

Addition of two or more such waves with different phases and amplitudes is a difficult trigonometric problem. It becomes almost unrealistic considering the huge numbers of waves scattered from all the electrons within a protein crystal. However, the ability to represent waves as vectors comes to the rescue as it allows simplifying the process considerably. In a vector representation of a wave, its amplitude is the length of the vector and the angle of the vector with the horizontal axis (x-axis) is its phase. The addition of two waves as vectors is depicted in Figure 3-3a.

Waves can also be represented as vectors in a complex plane (Argand diagram). A vector in a complex plane can be broken down into a real (projection on x-axis) and an imaginary part (projection on y-axis) and is expressed as a sum of the two (Fig. 3-3b). The real and the imaginary components of different waves can be added separately, thus, simplifying the overall process. In addition, a vector in a complex plane can also be interpreted as an exponential function (Equation 3.4), which simplifies the expression of complex equations.

If z is a vector in the complex plane, it can be broken down into a real and an imaginary component as follows:

$$\mathbf{z} = z(\cos\alpha + i\sin\alpha) = z\exp[i\alpha] \tag{3.4}$$

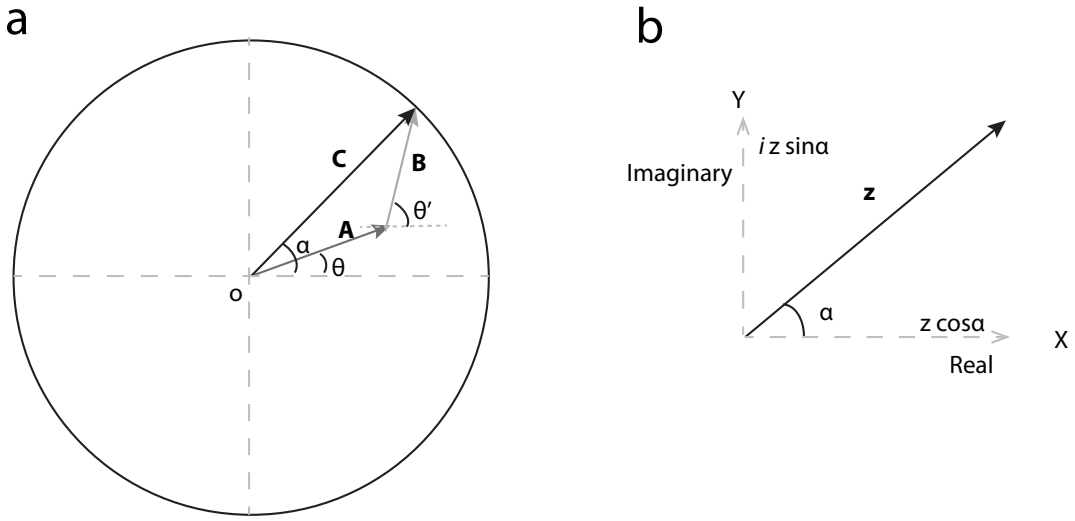


Figure 3-3: Representation of vectors

(a) Addition of vectors **A** and **B** to give the resultant vector **C**. The respective phases are shown. The phase of the resultant vector is different from both the component vectors. (b) A vector can be depicted as a complex entity in an Argand plane. Here the vector **z** can be broken down into a real component along the horizontal axis (real) axis and an imaginary component along the vertical (imaginary) axis.

3.4 Diffraction and the Bragg's Law

The diffraction pattern observed is a result of interference of the waves scattered in different directions from the electrons. In most of the directions, the waves cancel each other out and this is known as destructive interference. However, in some of the directions, the waves are scattered in phase (or approximately in phase) resulting in constructive interference and intensity maxima being observed. The condition for scattered waves reaching a wavefront in phase resulting in constructive interference was first worked out by Lawrence Bragg and is known as Bragg's Law

Bragg first showed that lattice points could scatter in phase only if the angle of incidence and the scattering angle of the waves are the same. Thus, the plane containing these points can be considered as a mirror plane that can reflect the incident X-rays. Waves reflected from the same plane travel exactly the same path-length, and therefore reach a scattered wavefront in phase. This principle was then extended to explain how scattering from more than one equivalent lattice planes could result in constructive interference.

As can be seen from Figure 3-4, the path difference between two waves reflected from two equivalent lattice planes is equal to 2dsinθ and based on the condition for constructive interference, this should be an integral multiple of the wavelength (λ) for the waves to reach the diffracted wavefront in phase. This is known as the Bragg's law and can be written in the form of Equation 3.5

$$n\lambda = 2d\sin\theta \tag{3.5}$$

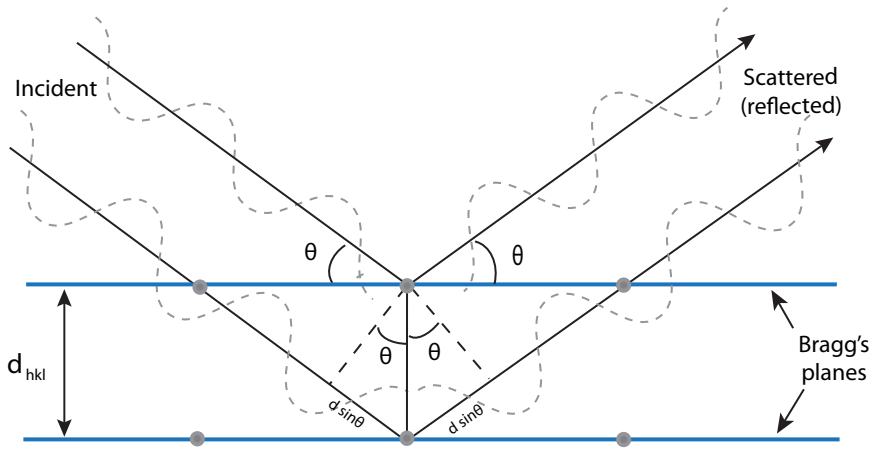


Figure 3-4: Braggs law

Diffracted waves undergo constructive interference only if their path difference is an integral multiple of the wavelength (λ). Here the two waves have a path difference of 1λ and the condition for constructive interference is satisfied. The lattice planes in this case are equivalent and can be identified by the same set of *hkl* values (Miller indices).

As a set of parallel lattice planes in a crystal can be assigned the Miller indices (h k l), Equation 3.6 can also be written as:

$$n\lambda = 2d_{hkl}\sin\theta \tag{3.6}$$

d_{hkl} being the distance between the planes defined by the indices (hkl).

Equation 3.6 can be rearranged as:

$$d_{hkl} = \frac{n\lambda}{2}\left(\frac{1}{\sin\theta}\right) \tag{3.7}$$

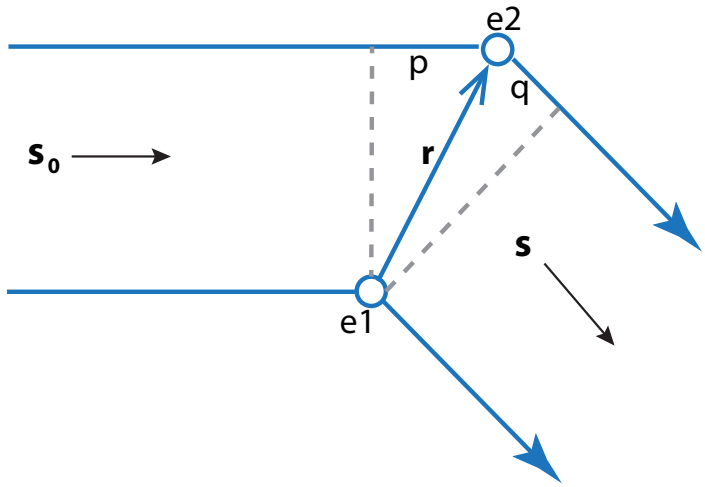
So, the distance between the lattice planes is inversely proportional to the glancing angle. As a consequence, closer the spacing between equivalent lattice planes, higher should be the glancing angle for the diffracted waves to interfere constructively.

3.5 Obtaining electron density from diffraction data

This section provides a mathematical description of diffraction and shows in a stepwise manner how the electron density is obtained from the diffraction data.

3.5.1 Mathematics of diffraction

The relative phase of a diffracted wave can be expressed mathematically using a system of two electrons.



In this system, $\mathbf{s}_o$ represents the incident wave vector and **s** the scattered wave vector, both having a magnitude of $1/\lambda$. The position of electron e2 with respect to electron e1

is **r** and the path difference of the wave through the electrons e2 compared to that of the electron e1 is p + q. The phase of the wave scattered by e2 is given by the expression:

$$2\pi(s - s_0).r \tag{3.8}$$

and substituting (**s** - $\mathbf{s}_0$) = **S**, the phase can be expressed as:

$$2\pi S.r \tag{3.9}$$

S is known as the diffraction vector, which is perpendicular to the Bragg planes and has a magnitude equal to the reciprocal of the spacing between a pair of Bragg planes (1/d). Thus, in mathematical terms, the diffraction vector (**S**) defines the reciprocal space. The diffraction vector for a set of planes in the reciprocal space defined by the indices (hkl) can be represented as:

$$S = h\,\mathbf{a}^* + k\,\mathbf{b}^* + l\,\mathbf{c}^* \tag{3.10}$$

where a*, b* and c* define the reciprocal unit cell axes.

The expression in Equation 3.9 shows that the phase at a particular position depends on the position vector **r.** Thus, the phase at another position of the same electron or of another electron located at a different position (say at **r** + **R** from origin) would result in a shift of phase compared to that in the original position and would be proportional to (**r** + **R**).

The real space position vector **r,** with fractional coordinates (x, y, z) can be also expressed as fractions of the unit cell axes a, b and c such that:

$$\mathbf{r} = x\,\mathbf{a} + y\,\mathbf{b} + z\,\mathbf{c} \tag{3.11}$$

Substituting the values of S and r, the expression **S.r** can be written as follows:

$\mathbf{S}.\mathbf{r} = (h\,\mathbf{a}^* + k\,\mathbf{b}^* + l\,\mathbf{c}^*) \,.\, (x\,\mathbf{a} + y\,\mathbf{b} + z\,\mathbf{c})$

$$\mathbf{S}.\mathbf{r} = h\,x + k\,y + l\,z \qquad (3.12)$$

(since $\mathbf{a}.\mathbf{a}^* = \mathbf{b}.\mathbf{b}^* = \mathbf{c}.\mathbf{c}^* = 1$)

Thus, (hx + ky + lz) can be substituted for $\mathbf{S}.\mathbf{r}$ in equations involving this term.

3.5.2 Structure factor

X-rays are scattered by the centrosymmetric electron cloud of an atom. The electron densities $\rho(\mathbf{r})$ and $\rho(-\mathbf{r})$ at positions $\mathbf{r}$ and $-\mathbf{r}$ respectively are thus the same. The scattering factor (f) at the position $\mathbf{r}$ of an atom can be expressed as:

$$f = \int_{r} \rho(\mathbf{r}) \exp[2\pi i \mathbf{r}.S] d\mathbf{r} \qquad (3.13)$$

If the contribution of the electron cloud at position $(-\mathbf{r})$ is also considered, then the imaginary components of the scattering vectors of the centrosymmetric electron cloud, which have equal magnitude but opposite direction, cancel out. So, scattering by an atom is only composed of the real components and can be written as:

$$f = 2\int_{r} \rho(\mathbf{r}) \cos[2\pi \mathbf{r}.S] d\mathbf{r} \qquad (3.14)$$

Scattering from a unit cell can be thought of as a sum of the scattering from its individual atoms. The scattering arising from a unit cell depends on the arrangement of the atoms and their electron densities, which in turn is determined by the symmetry and is therefore known as the structure factor. A structure factor describes the contributions of all the atoms of a unit cell toward the formation of a single diffraction spot on the detector. Thus, in a unit cell containing electrons at positions $\mathbf{r}_j$ (j=1, 2, 3, ..., n) with respect to the origin, the structure factor [$\mathbf{F}(\mathbf{S})$] can be written as a sum of the atomic scattering factors (f_j):

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j.\mathbf{S}] \tag{3.15}$$

The structure factor equation can also be written in terms of the fractional coordinates, using Equation 3.13, as:

$$\mathbf{F}(hkl) = \sum_{j=1}^{n} f_j \exp[2\pi i(\mathrm{h}x_j + ky_j + lz_j)] \tag{3.16}$$

3.5.3 The temperature factor

The temperature factor, also known as the B-factor, describes the decrease in the measured intensity as a function of resolution. Atoms undergo vibrations in a temperature-dependent manner resulting in a dynamic disorder in the protein crystals. This leads to a smearing effect, the extent of which varies across the unit cells. As a result, the X-rays hit slightly shifted atoms at identical positions in different unit cells, causing a reduction in the scattered intensity with the effect increasing with higher resolution. The difference in scattering intensities due to the smearing is accounted for by multiplying the atomic scattering factors with a temperature-dependent factor. Given the limited resolution that can be achieved for a protein crystal, the temperature factor is assumed to be isotropic and is expressed as:

$$T_{(iso)} = \exp[-B\frac{\sin^2\theta}{\lambda^2}] \tag{3.17}$$

and using the relation $\sin\theta/\lambda = 1/2d$ (Bragg's law), Equation 3.17 can be written as:

$$T_{(iso)} = \exp[-\frac{B}{2d^2}] \tag{3.18}$$

The temperature factor, otherwise known as the Debye-Waller factor, also accounts for any static disorder due to crystal defects apart from the dynamic disorder. The quantity B represents the breadth of smearing and is expressed as:

$$B = 8\pi^2 u^2 \tag{3.19}$$

where u is the root mean square displacement.

Taking the B-factor into consideration, the structure factor equation (3.16) can be modified as:

$$\mathbf{F}(hkl) = \sum_{j=1}^{n} f_j \exp[2\pi i(\mathrm{h}x_j + ky_j + lz_j)]\exp[-B_j \sin^2\theta/\lambda^2] \tag{3.20}$$

3.5.4 Fourier analysis

Complicated periodic functions can be represented as a sum of a series of sine and cosine functions. This was shown by Joseph Fourier and such a series whose sum represents a complicated function is known as a Fourier series. A Fourier series can thus be used to describe a complicated wave as a sum of a series of simple waves whose frequencies, known as harmonics, are integral multiples of a fundamental frequency. Likewise, Fourier analysis can be used for any periodic function including the complex structure factors and the electron density distribution within a unit cell.

An important application of Fourier analysis in X-ray crystallography is in the reconstruction of the electron density from the calculated amplitudes and phases. A general explanation of Fourier transformation is given below.

If the complex function $F(h)$ can be written as a Fourier series of the function $f(x)$, then:

$$F(h) = \int_{\infty}^{-\infty} f(x) e^{2\pi i(hx)}\, dx \tag{3.21}$$

$F(h)$ is called the Fourier transform of the function $f(x)$ and the units of the variables h and x are reciprocal of each other. Since Fourier transformation is reversible, the equation can be rearranged as:

$$f(x) = \int_{\infty}^{-\infty} F(h) e^{-2\pi i(hx)}\, dx \tag{3.22}$$

Thus, the function $f(x)$ can be described as a reverse Fourier transform of the function $F(h)$.

3.5.5 Calculation of the electron density

The structure factor, as described before, is a function of the electron density distribution of the atoms in a unit cell. Based on Equation 3.15, the structure factor can be expressed as a sum of the individual atomic scattering factors. However, it can also be thought of as an integral of the density distribution over the entire unit cell volume (V):

$$\mathbf{F}(\mathbf{S}) = \int_{\substack{\text{unit}\\ \text{cell}}} \rho(\mathbf{r}) \exp[2\pi i \mathbf{r}_j.\mathbf{S}]\, dv \tag{3.23}$$

where $\rho(\mathbf{r})$ is the electron density at position **r** in the unit cell.

If x, y and z are the fractional coordinates at position **r**, and V is the volume of the unit cell, and r.S is replaced with (hx + ky + lz) (Equation 3.12), then **F**(**S**) is equivalent to **F**(hkl) and can be expressed as an integral of the infinitesimally small density elements defined by ρ(xyz) and can be written as:

$$\mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz \tag{3.24}$$

Equation 3.24 shows how to obtain the amplitude and phases (which define the structure factor, **F**(hkl) or **F**(**S**)) from the electron density. However, the reverse operation, that is, obtaining the electron density distribution from the calculated amplitudes and phases, is also possible and is achieved by the mathematical operation called Fourier transformation (as described in the previous section).

F(hkl) and ρ(xyz) are the Fourier transforms of each other and therefore Equation 3.24 can be rearranged as:

$$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\mathbf{F}(hkl)\exp[-2\pi i(hx+ky+lz)] \tag{3.25}$$

The integration in Equation 3.24 is replaced by summation because Equation 3.25 involves **F**(hkl) or the structure factors, each of which describe a discrete spot in the diffraction pattern rather than a continuous function like density distribution.

3.6 The phase problem

As explained above, the electron density is a Fourier series with the structure factors as coefficients. Each structure factor term describes a single reflection and is defined by amplitude as well as phase. The amplitudes of the structure factors can be calculated from the intensity of the reflections; however, the relative phases of the diffracted waves that give rise to the reflections cannot be directly obtained from the diffraction data and is known as the phase problem. Thus, in order to reconstruct the electron density from the diffraction data, the relative phases need to be calculated using one, or sometimes a combination of more than one, of the three prevalent phasing methods. Two of these depend on heavy atoms, which act as markers for *de novo* calculation of the phases. These include the isomorphous replacement and the anomalous scattering methods. The third method to obtain the phases is known as molecular replacement and depends on the prior availability of homologous structures that could be used for extracting the phases of a protein with a similar sequence.

3.6.1 The Patterson function

Calculation of phases by the heavy atom methods depends on locating the heavy atoms in the unit cell of the derivatised crystals. Obtaining the heavy atom positions is achieved by calculating the Patterson function, a variation of the electron density function. It consists of a Fourier series with square of amplitudes (or intensities) as

coefficients but without phases. Thus, the Patterson function, unlike the electron density, can be directly calculated from intensities of the diffraction data and does not depend on the phases. The Patterson function can therefore be described as:

$$P(uvw) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|\mathbf{F}(hkl)\right|^2 \cos[2\pi i(hu + kv + lw)] \tag{3.26}$$

where u, v and w are the coordinates in the Patterson space and can be replaced with the positional vector **u** instead.

The Patterson function can also be written as:

$$P(\mathbf{u}) = \int_{\mathbf{r}} \rho(\mathbf{r}) \times \rho(\mathbf{r} + \mathbf{u})\, dv \tag{3.27}$$

According to this equation, the Patterson function can be obtained by integrating the densities at position **r** and (**r** + **u**) over the entire Patterson space. The function is only significant if the corresponding densities have non-zero values and thus represent two atoms in the real unit cell. Hence, the Patterson coordinates actually represent the relative positions of the atoms under consideration and manifest themselves as peaks corresponding to interatomic vectors. Such peaks are located within a Patterson unit cell, the dimensions of which are the same as the unit cells of the crystal. The Patterson map of a molecule with N atoms will have N^2 number of peaks, out of which N peaks relate to self-vectors of the atoms and thus, practically, the total number of peaks observed will be N^2-N. An example of the construction of a Patterson unit cell for a molecule with three atoms has been shown in Figure 3-5.

The Patterson map of a protein will therefore contain a huge number of peaks and making sense of such a map is impossible, especially considering the associated noise that often masks genuine peaks. However, since the size of the peaks in a Patterson map is proportional to the product of the atomic numbers of the pairs of atoms involved, the peaks corresponding to the difference vectors of the heavy atom positions can be easily distinguished from peaks involving the lighter protein atoms. From the heavy atom Patterson map, the number of heavy atoms present in a unit cell can be calculated and

then their actual positions are found out using a trial method. In cases where molecules within the unit cell are related by symmetry, the Patterson peaks corresponding to the interatomic vectors of equivalent atoms, are concentrated at specific locations, which are known as Harker planes or Harker lines. This observation provides the idea as to where to look for the peaks in the Patterson space and thus, simplifies the problem of calculating the coordinates.

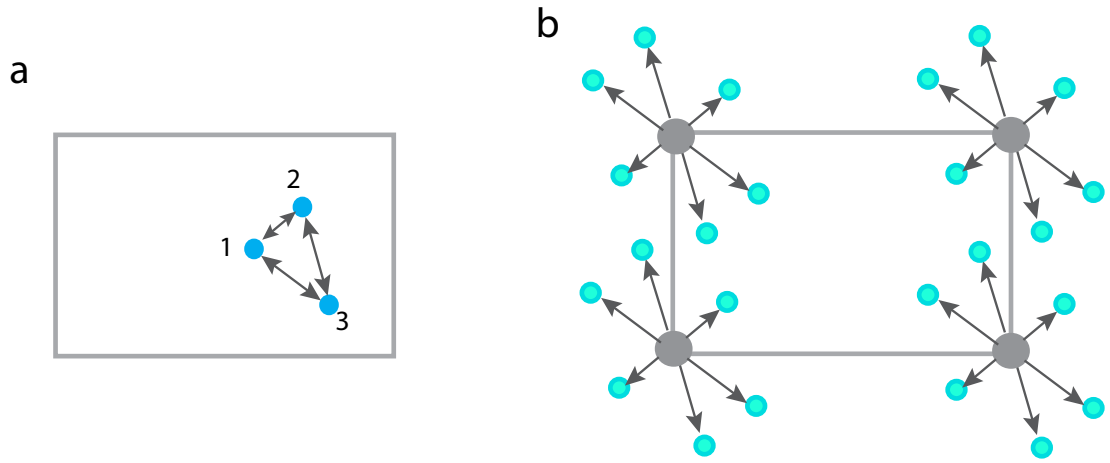


Figure 3-5: Construction of a simple Patterson unit cell

(a) a real space unit cell with three atoms; the interatomic vectors are shown by arrows and (b) a Patterson unit cell with the Patterson peaks (interatomic vectors in cyan and self-vectors in grey); a pair of vectors of equal magnitude and opposite direction are possible between each pair of atoms making the Patterson unit cell centrosymmetric. The vertices of the Patterson cell correspond to the self-vectors and therefore have strongest peaks.

3.6.2 Isomorphous replacement method

Isomorphous replacement method involves addition of one or a few heavy atoms (strong scatterers) into a protein crystal so that they cause a measurable difference in the diffraction intensities. Diffraction data from the native (only protein) and derivative crystals are collected so that the intensities of equivalent spots from both the datasets can be compared. Thus, one of the most important criteria for successful estimation of phases using isomorphous replacement method is isomorphism between the native and derivative crystals.

The method is based on the idea that the scattering from a derivative crystal containing both protein and heavy atoms is a sum of the scattering by the protein and heavy atoms. This can be represented by the vector equation.

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H \tag{3.28}$$

where $\mathbf{F}_{PH}$ is the scattering by the derivative, $\mathbf{F}_P$ by the protein alone and $\mathbf{F}_H$ by the heavy atoms.

For a particular reflection, the square of the difference of its amplitudes with and without the heavy atom contribution, $\left[|F_{PH}|-|F_P|\right]^2$, is equal to the contribution of the heavy atoms toward its intensity, $|F_H|^2$. Thus, the difference in the intensities in the diffraction data from the derivative and the native protein crystals is used to calculate the Patterson function, from which the heavy atom coordinates can be calculated as described before. The phases of the protein ($\mathbf{F}_P$) can be calculated based on Harker diagrams shown in Figure 3-6 When a single heavy atom derivative is used (single isomorphous replacement; SIR), it usually leads to a phase ambiguity. The problem is resolved by collecting additional derivative data and this is known as multiple isomorphous replacement.

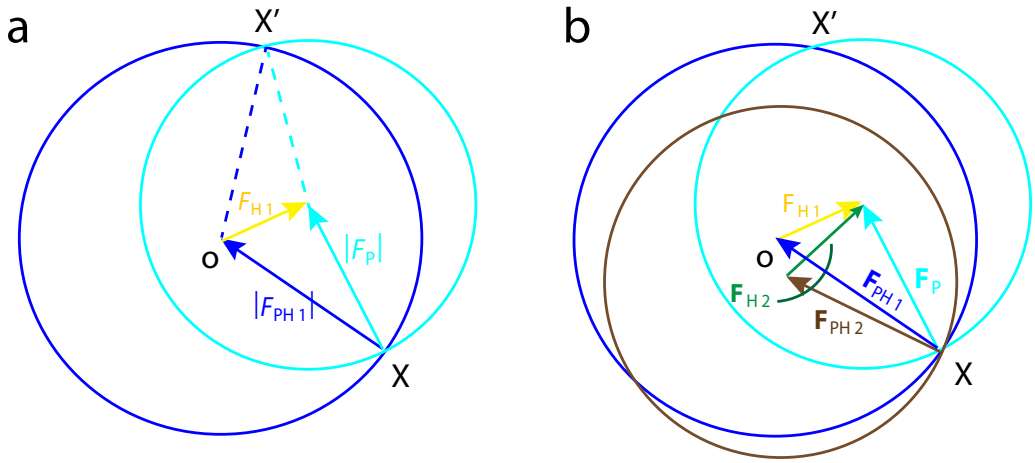


Figure 3-6: Phase estimation by isomorphous replacement method

Harker diagrams for (a) SIR and (b) MIR. The circle with centre at O is drawn with a radius equal to magnitude of $\mathbf{F}_{PH1}$. Thus the vector $\mathbf{F}_{PH1}$ can lie anywhere on this circle. At O, the vector $\mathbf{F}_{H1}$ is drawn (obtained from the difference Patterson). Taking the end of the vector $\mathbf{F}_{H1}$ as the centre, a second circle with a radius equal to the magnitude of $\mathbf{F}_P$ is drawn. The points of intersection of the two circles (X and X') represent the possible phases of the vector $\mathbf{F}_P$, resulting in an ambiguity. This ambiguity in phase determination is resolved by drawing circle(s) corresponding to the scattering vector of additional derivatives (shown as $\mathbf{F}_{PH2}$ in the Harker diagram on the right). Adding additional derivative data thus helps in obtaining a single solution for the phase of the protein.

3.6.3 Anomalous scattering method

Diffraction, under normal conditions, results when the electrons absorb energy from the incident radiation, are set to vibrate and re-emit radiation of the same frequency. This is called coherent scattering and can occur only when electrons can vibrate freely inside atoms. Under such conditions, reflections related by inversion through the origin (Friedel pair), have the same amplitude but opposite phase, known as Friedel's law (Equation 3.30). However, the electrons absorb differently as the wavelength of the incident radiation varies. At certain wavelengths, characteristic of the type of atom and the quantum level of the electron (K, L, M, etc.), the absorption decreases abruptly and the corresponding region of wavelength is called the absorption edge. At wavelengths just short of the absorption edge of an element, which correspond to atomic transitions, the energy of the X-rays is sufficient to eject electrons out of their shell (transition energy). Once ejected, the electrons do not behave as free electrons any more and scatter radiation with a frequency and phase different from the incident wave (incoherent scattering). Under these abnormal or anomalous conditions, Friedel's law does not hold and the phenomenon is known as anomalous scattering. It was first shown by Bijvoet that these differences in the intensities of Friedel pairs, arising due to the presence of anomalous scatterers, could be used to estimate the phase of the diffracted waves.

$$\left|F_{hkl}\right| = \left|F_{-h-k-l}\right| \tag{3.29}$$

At X-ray wavelengths generally used for diffraction experiments, lighter atoms like carbon, oxygen or nitrogen do not show anomalous scattering but elements like sulphur or those heavier do. These atoms are known as heavy atoms and one of the elements widely used for anomalous scattering experiments is selenium. Selenium can be incorporated into a protein by growing *E. coli* cells expressing the target protein in a minimal medium containing selenomethionine instead of normal methionine. Anomalous data can be collected by adjusting the wavelength of the X-rays according to the heavy atom used, which can easily be achieved in the synchrotron sources.

When the heavy atoms present in a protein absorb X-rays with a wavelength near its absorption edge, they emit radiation with an altered phase. The scattering factor, under these conditions, is composed of an anomalous component apart from the normal component (the scattering factor at wavelengths far apart from the edge). Furthermore, the anomalous component of the altered atomic scattering factor is made of a real part and an imaginary part that is perpendicular to the real part. This can be represented as in Equation 3.30 and is depicted in Figure 3-7.

$$f_{anom} = f_N + \delta f' + f'' \tag{3.30}$$

Replacing the real components ($f_N + \delta f'$) by f', Equation 3.30 can be written as:

$$f_{anom} = f' + f'' \tag{3.31}$$

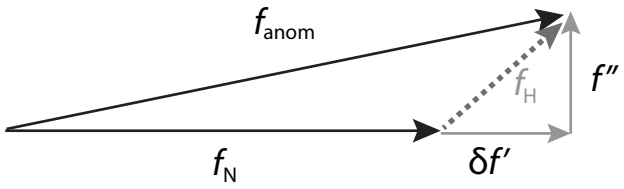


Figure 3-7: Vector representation of anomalous scattering

f_N is the normal atomic scattering factor at wavelengths far apart from the absorption edge. The anomalous component (f_H) consists of a real part ($\delta f'$) and an imaginary part (f''), perpendicular to each other. The total anomalous scattering is denoted by f_{anom}.

3.6.3.1 *Relation between Friedel mates during anomalous scattering*

As was mentioned earlier, as a result of anomalous scattering, Friedel's law is broken and thus, the Friedel mates have different intensities and are no longer related by opposite phases to each other. The Friedel pair of reflections that differ from each other due to anomalous scattering are known as Bijvoet pairs. This has been depicted in Figure 3-8, where the $\mathbf{F}_{HP}^{+}$ and $\mathbf{F}_{HP}^{-}$ represent the anomalous scattering factors of a

Bijvoet pair and are oriented at different angles from the horizontal axis. The real contributions to the anomalous scattering, $\mathbf{F}_{H(r)}^{+}$ and $\mathbf{F}_{H(r)}^{-}$ are still the reflections of each other along the X-axis and are thus, related by opposite phases. However, the imaginary parts, $\mathbf{F}_{H(i)}^{+}$ and $\mathbf{F}_{H(i)}^{-}$ are inverted reflections of each other (rather than ordinary reflections) and this difference is responsible for the differences in the intensity as well as the phase of Friedel mates under conditions of anomalous scattering.

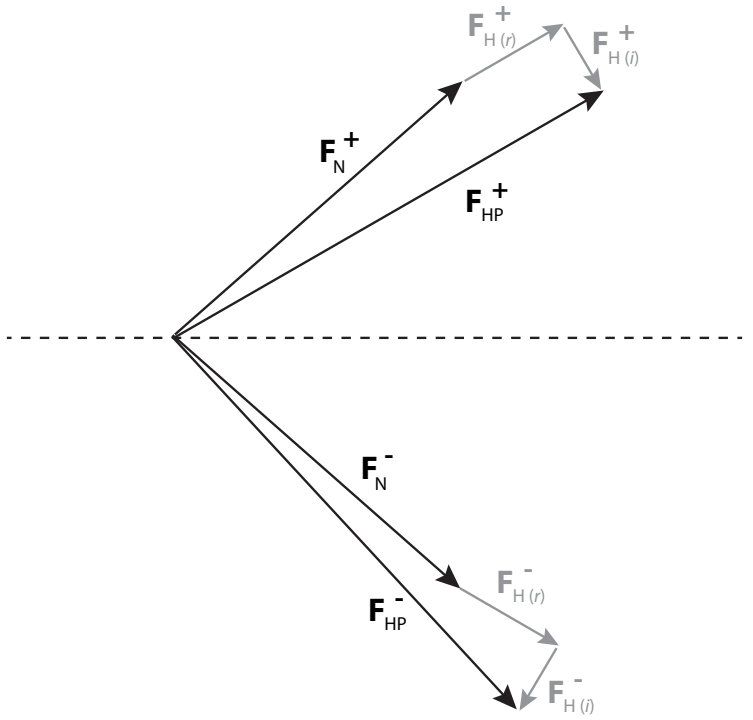


Figure 3-8: Violation of Friedel's law

Friedel's law does not relate the scattering factors $F_{HP}+$ and $F_{HP}-$ as they are not the mirror images of each other, unlike F_N+ and F_N-. The corresponding pair of reflections, under the conditions of anomalous scattering is known as a Bijvoet pair.

3.6.3.2 *Calculation of phases from the anomalous data*

From Figure 3-8 and Equation 3.30,

$$\mathbf{F}_{HP}^{+} = \mathbf{F}_{N}^{+} + \mathbf{F}_{H(r)}^{+} + \mathbf{F}_{H(i)}^{+} \qquad (3.32)$$

where $\mathbf{F}_{HP}$ is the anomalous scattering factor, $\mathbf{F}_N$ is the non-anomalous or normal scattering factor and $\mathbf{F}_{H(r)}$ and $F_{H(i)}$ are the real and imaginary parts of the anomalous scattering. All the terms in Equation 3.32 involve the reflection with a positive phase and the same equation holds for its Friedel mate (with a negative phase). Solving Equation 3.32 for only $\mathbf{F}_N^+$ results in a phase ambiguity, as in the case of single isomorphous replacement. However, the actual phase can be obtained when the anomalous scattering component of the Friedel mate ($\mathbf{F}_N^-$ in this case) is taken into account. This explains the importance of anomalous scattering in calculating the phases of the diffraction data.

Equation 3.33 can be rearranged as (without specifying any Friedel mate):

$$\mathbf{F}_N = \mathbf{F}_{HP} - \mathbf{F}_{H(r)} - \mathbf{F}_{H(i)} \qquad (3.33)$$

The magnitudes of $\mathbf{F}_{H(r)}$ and $\mathbf{F}_{H(i)}$ are constants for a given element (e.g., selenium) while their phases can be calculated from the positions of the heavy atoms (calculated by Patterson methods). Thus, $\mathbf{F}_N$ can be obtained as all the terms on the right hand side of Equation 3.33 can be calculated with the help of the heavy atom anomalous data.

Solution of the phase problem by MAD or multiwavelength anomalous diffraction method involves collection of anomalous data at different wavelengths. The data are generally collected at wavelengths corresponding to – (a) the peak, where the anomalous signal is maximum; (b) the absorption edge, also known as the inflection, and (c) the remote, which is far from the peak wavelength. The real and imaginary parts of the anomalous scattering, $\mathbf{F}_{H(r)}$ and $\mathbf{F}_{H(i)}$ vary considerably at different wavelengths and this difference is used for precise calculation of the phases. Besides, the intensities of the reflections also vary at different wavelengths and these differences can also be harnessed to aid in determination of the phase.

3.6.4 Molecular replacement

Phasing by molecular replacement (MR) depends on prior availability of structures of proteins homologous to the target protein sequence. Proteins with adequate sequence homology generally have a similar three-dimensional arrangement of their Cα backbones, allowing the phase information of the known structure to be borrowed in order to calculate the phases of the unknown target structure. MR thus involves a relatively direct approach compared to the indirect phasing methods based on heavy atom detection that were discussed earlier. However, in order to use the available phase information, the relative orientation as well as the location of the template and the target molecules within their respective unit cells need to be determined. Thus, the problem can be solved in two steps; first one being a rotation to bring the molecules in the same orientation and the next is to determine the displacement required to place an identically oriented molecule onto the other. In three-dimensions, a rotation can be fully described by three angles about the respective axes, while a translation can similarly be described by three displacements (along the three axes). Thus, six parameters need to be varied if the relative position of two similar molecules is to be determined. To simplify the process, in practice, molecular replacement is broken down into a rotation function search followed by a translation function search. The Patterson function, as with the other methods of phase determination, is an integral part of phase determination by MR and it allows for separate determination of the rotation and the translation functions.

3.6.4.1 *The rotation function*

The Patterson function of a single molecule represents only its interatomic vectors called self-Patterson vectors and a rotation in the molecule results in a rotation of the corresponding Patterson vectors. Thus, in this simplified case, the relative orientation of two molecules can easily be determined by comparing their Patterson functions. However, in crystals, the molecules are related by symmetry operations and therefore, the Patterson function in such cases also consist of vectors between atoms of different molecules, called cross-Patterson vectors. A rotation in a molecule can be defined entirely by rotating its self-Patterson vectors, so, the cross-Patterson vectors, which are

generally longer in magnitude, are not considered for calculating the rotation function. Practically, this is achieved by restricting the rotation search to a radius just smaller than the dimensions of a molecule. Generally, for spherical molecules the search radius is 75% of the diameter whereas for elongated ones, an average of the three semi-axes is used as a guide.

So, based on the above description, the Patterson functions representing specific orientations of the model need to be sampled over the entire space covered by the self-Patterson vectors of the target in order to analyse the corresponding fits between the rotated model and the target. If a rotation operation, C, is carried out on the model such that its initial and the rotated Patterson functions are given as $P(\mathbf{u})_M$ and $P(C\mathbf{u})_M$ while the Patterson function of the target is denoted as $P(\mathbf{u})_T$, then the rotation function, R, for the operation can be expressed as:

$$R(C) = \int_U P(C\mathbf{u})_M \times P(\mathbf{u})_T \, d\mathbf{u} \tag{3.34}$$

where U is the volume defined by the self-Patterson vectors.

The rotation function R is calculated at a number of positions **u** over the entire space U and has a maximum value when there is an overlap between the respective Patterson maps and is thus also known as the overlap function. The significance of a particular operation, C, is evaluated by comparing the corresponding rotation function with the mean of R(C) values obtained over the entire Patterson space U.

3.6.4.2 *The translation function*

Identification of the correct rotation operation is followed by determination of the translation function such that the observed intensities of the target match the calculated Patterson function of the virtually shifted model. The translation search aims to find the position of the target molecule within a unit cell in which the molecules are related by some sort of symmetry operation(s). If there is no symmetry within the unit cell, as in

the case of the space group P1, the choice of origin is arbitrary and therefore there is no need to define the translation function. However, the contents of the unit cell are generally related by symmetry operations and the translation function is used to fix the origin of the unit cell as well as locate the symmetry operator with respect to the correctly oriented molecule. Identification of the correct translation makes use of the cross-Patterson vectors between the symmetry related molecules. The volume of a translation search is known as a Cheshire cell, a unique volume within the unit cell that varies according to the choice of origin for a specific space group.

3.7 Refinement of protein structures

Once the phase problem is solved and an electron density map is calculated, an initial atomic model of the protein can be obtained. However, the initial model generally resembles the best possible model, given the observed intensities, only approximately. Structure refinement aims to narrow down this gap and works on the principle that the calculated (i.e., the model) and the observed structure factors should converge as much as possible to obtain the best possible model from the data. The agreement between the observed and the calculated structure factors can be described by the R-factor (refinement parameter) and can be expressed as in Equation 3.35.

$$R = \frac{\sum_{hkl} \left\| F_{\text{obs}} \right| - \left| F_{\text{calc}} \right\|}{\sum_{hkl} \left| F_{\text{obs}} \right|} \tag{3.35}$$

The model structure factors depend on the calculated atomic coordinates as well as the B-factors (Equation 3.19) and therefore each atom is associated with four variable parameters (x, y, z coordinates and the B-factor). These parameters associated with each atom in the model need to be varied so that a global minimum can be achieved, which corresponds to the maximum possible agreement between the calculated and observed structure factors. A factor important for satisfactorily refining these variable parameters and consequently reaching the global minimum is to have on disposal sufficient number of observations and as a rule of thumb, the number of observations should be at least

three times the total number of refinable parameters. For example, for a protein with X number of non-hydrogen atoms, which will consequently have 4X refinable parameters, a minimum of 3 times of 4X, that is, 12X unique reflections need to be recorded.

While refining the parameters, the stereochemistry of small molecules, like peptides, is used as a guide to define the values of bond lengths and angles and may be enforced in two possible ways – either as constraints or restraints. When the atoms are considered as rigid groups and only one or a limited number of parameters, e.g., the torsion angle of the peptide bond, are allowed to be varied at a time, then the refinement is said to be constrained. Constraints, therefore, have an effect of decreasing the number of parameters that need to be varied during the refinement process. On the other hand, during a restrained refinement, the atoms are considered to be relatively independent and the associated parameters are allowed to vary within a limit. As a result, enforcing restraints do not cause any decrease in the number of parameters that need to be refined, but might still prove to be a valuable tool as it leads to a more reasonable assignment of values to the parameters.

3.7.1 Least squares refinement

The least-squares minimisation is a reciprocal space refinement technique initially applied for small-molecules. The principal could also be used with relative success for refining protein structures and a least-squares protocol with constrained parameters was especially useful for refining data with lower observation to parameter ratios. Refinement by least squares method aims to minimise a function Q, which is a weighted (denoted by the term '*w*') sum of the deviation between squares of observed and calculated structure factor amplitudes (Equation 3.36).

$$Q = \sum_{hkl} w\left[\left|F_{obs}\right| - k\left|F_{calc}\right|\right]^2 \tag{3.36}$$

here '*w*' is a weighting term equal to the inverse of squared standard deviation, σ and '*k*' is a scale factor.

The limited success of the least-squares method can be attributed to some of the basic assumptions on which the method is based. During least squares refinement it is assumed that the probability distributions of observed and calculated structure factor amplitudes are Gaussian functions and that the latter is independent of model parameters (so they can be constrained); also, the observed phases are assumed to be correct. However, for protein structures these assumptions do not always hold true and as a result, the structures refined using this method showed poor refinement statistics and were also biased toward the model.

3.7.2 Maximum likelihood refinement

The maximum-likelihood method is a more general form of the least squares method and unlike the latter, it does not assume the structure factors to be Gaussian distributions only. The method is based on the principle that given a model (that has been calculated), what is the probability that the observations would have been made; thus, higher the probability, higher is the likelihood or agreement between the model and the data. The aim of a refinement process, however, is to obtain the probability of a model, given the data (denoted as p[model; data]). These two probabilities (p[data; model] and p[model; data]) are related by Bayes' theorem as follows:

$$p[\text{model};\text{data}] = \frac{p[\text{model}] \;.\; p[\text{data};\text{model}]}{p[\text{data}]} \tag{3.37}$$

Since the data does not vary, the denominator is a normalisation constant and can be omitted. The term p[model] is significant only when there is a prior information on the model so it can be neglected as well and so, Equation 3.37 can be written as:

$$p[\text{model};\text{data}] = p[\text{data};\text{model}] \tag{3.38}$$

The term 'data' can be replaced with F_{obs} and 'model' with $\mathbf{F}_{calc}$. Since the normalisation term is omitted, the right hand side expression of Equation 3.38 is called likelihood (L) instead of probability. In describing the likelihood function, the

observations are considered independent of each other. The total likelihood of the model is described as a negative sum of the log of likelihood (Equation 3.39).

$$-\log L_{\text{total}} = -\sum \log\left[\mathrm{p}\left|F_{\text{obs}}\right|; p\left(\mathbf{F}_{\text{calc}}\right)\right] \tag{3.39}$$

Expressing likelihood in a logarithmic form allows it to be written as a sum, which is easier to interpret since a maximum of the function needs to be achieved for the optimal refinement of the model parameters.

The implementation of the maximum-likelihood method generally results in an overall superior quality of the model and phases compared to that following least squares refinement. Many of the popular refinement softwares used at present like REFMAC and PHENIX are based on the maximum-likelihood principle. The software SHARP, used for heavy atom refinement, also uses this principle for refining the heavy atom parameters.

Chapter 4. The crystal structure of WAPL domain

The N-terminal of the human Wpl1 orthologue was previously shown to be important for interaction with Pds5 (Kueng et al., 2006) but is highly variable both in terms of sequence length and composition and is also disordered based on secondary structure predictions. However, the function of the WAPL domain, which starts immediately after the N-terminal disordered region, was unknown although its conservation among eukaryotes suggests a key role of this domain. Therefore, crystal trials were carried out with a truncated Wpl1 consisting only of the WAPL domain. Crystal structure of this domain was elucidated and this chapter entails a detailed description of the crystallization experiments and structure solution process that was undertaken followed by an analysis of the structure.

4.1 Crystallisation

4.1.1 *S. cerevisiae* Wpl1$^{259-647}$ crystallisation

The WAPL domain stretches approximately from residue 259-647 in the *S. cerevisiae* Wpl1 (hereafter called ScWpl1$^{259-647}$) and this construct was used for the initial crystallization trials. The protein was purified as described earlier in Experimental methods section. The crystal screens were set-up as 400 nL sitting drops containing equal volumes of the well-solution and the protein at 10 mg/mL in 96-well plates, which were then incubated at 4 ˚C. An initial hit was obtained in the condition 0.1 M BisTris Propane pH 6.5, 0.2 M Na-citrate and 20% PEG 3350. The crystals observed in the screen were very small, polygonal shaped and needed further improvement in both size and quality. Optimization was performed using 24-well hanging drop plates in which 4 µL drops were set by adding equal proportions of well-solution and protein at 10 mg/mL concentration. After optimization, the condition in which bigger crystals were obtained was 0.1 M BisTris Propane pH 6.5, 0.2 M Na-citrate and 15% PEG 3350. These crystals were much bigger (30x30x10 µm) compared to the initial ones in the screen but when tested on an in-house X-ray source, diffracted only to about 8.5 Å (Figure 4-1). To further improve these crystals, a multivariate screen was performed, in which the protein (8 to 15 mg/mL) along with salt (0.1 to 0.5 M) and precipitant (10 to

25%) concentrations were varied together within a narrow range around the original concentrations. After this optimization, bigger crystals could be obtained when the salt and protein concentrations were slightly increased (0.3 M Na-citrate and 12 mg/mL protein) while keeping precipitant concentration constant (15% PEG 3350). The biggest crystals that could be obtained were of dimensions 50x50x15 µm but when tested, the

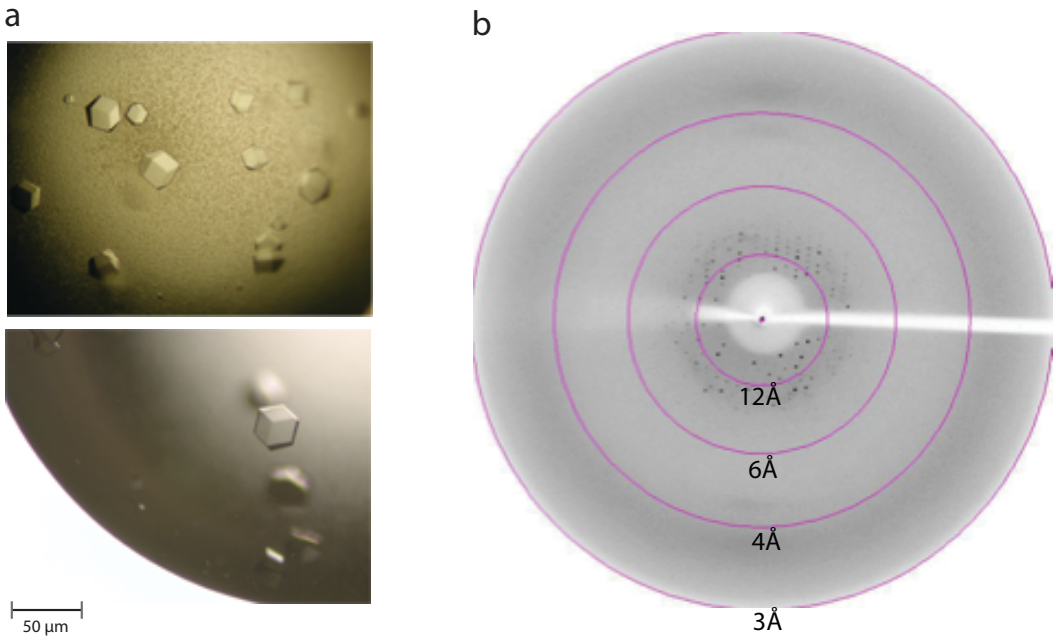


Figure 4-1: Crystals and diffraction of ScWpl1$^{259-647}$

diffraction only marginally improved to about 7 Å. Among other strategies that were tried were micro-seeding, dehydration, annealing and optimization of the cryoprotectant, none of which could adequately improve the diffraction of these crystals.

In order to get crystals belonging to a different crystal system, the following approaches were taken: (a) Change in the construct – several constructs starting from different nearby residues to the starting residue (E259) of the original construct were tried but these did not result in better crystals. (b) Different orthologues of Wpl1 – among the ones tried were human, *Drosophila* and *Ashbya* WAPL domains. The human and *Drosophila* WAPL domains failed to crystallize while the one from *Ashbya gossypii* did and is discussed in detail in the following section.

4.1.2 *A. gossypii* Wpl1$^{184-561}$ crystallisation

Ashbya gossypii is a filamentous fungus whose genome is similar to that of budding yeast with which it shares a common evolutionary origin. Around 95% of *A. gossypii* genes have orthologues in budding yeast and thus this organism provides a suitable alternative to *S. cerevisiae* for studying eukaryotic proteins. A sequence alignment of A. gossypii and S. cerevisiae Wpl1 proteins is shown in Figure 4-2.

In *A. gossypii*, the Wpl1 orthologue (AgWpl1) is slightly shorter in length, with 561 amino acid residues, compared to 647 in its budding yeast counterpart, with the WAPL domain extending from residues 184-561 (henceforth AgWpl1$^{184-561}$). The purified AgWpl1$^{184-561}$ was used to set up crystallization screens in 96-well sitting drop plates and 400 nL drops were set by adding 200 nL well-solution to an equal volume of the protein at 10 mg/ml concentration. Small crystals appeared within 2-4 days of setting up the trays in most of the screens across many different conditions. The crystals formed clusters of very small and thin plates stacked against each other through their flat surfaces (Figure 4-3). The best screen condition that produced crystals consistently when tested in 24-well plates using hanging drop method, was 0.02 M Na/K phosphate, 0.1 M Bis-tris propane pH 7.5, 20% PEG3350. The condition was further refined to 0.1 M Na/K phosphate, 0.1 M Bis-tris propane pH 7.5, 15% PEG 3350 and the protein concentration in the drops was reduced to 4 mg/mL to obtain bigger clusters. In order to obtain single crystals, several strategies were tried which included buffer and pH variation, screening for additives and microseeding. None of these techniques yielded desired results, so the individual plates had to be separated out carefully from the clusters to collect diffraction data. The Seleno-methionine (SeMet) - derivatised crystals were also grown in the same condition and formed clusters from which the individual plates had to be separated for mounting. The crystals were flash-frozen after soaking briefly in a solution containing 20% glycerol added to the mother liquor before mounting. The individual crystals diffracted to 2.8 to 3 Å in-house and about 2 Å in the synchrotron X-ray source (Figure 4-3). Both native and anomalous data were collected in the Diamond Light Source IO2 beamline. The structure solution process is discussed in the next section.

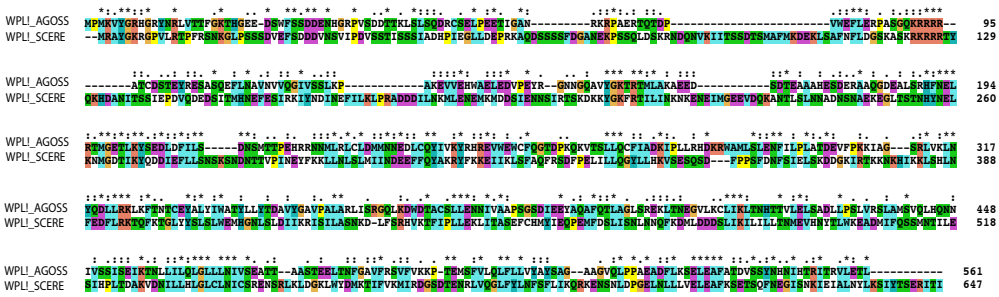


Figure 4-2: Alignment of Wpl1 sequences of *S. cerevisiae* and *A. gossypii*

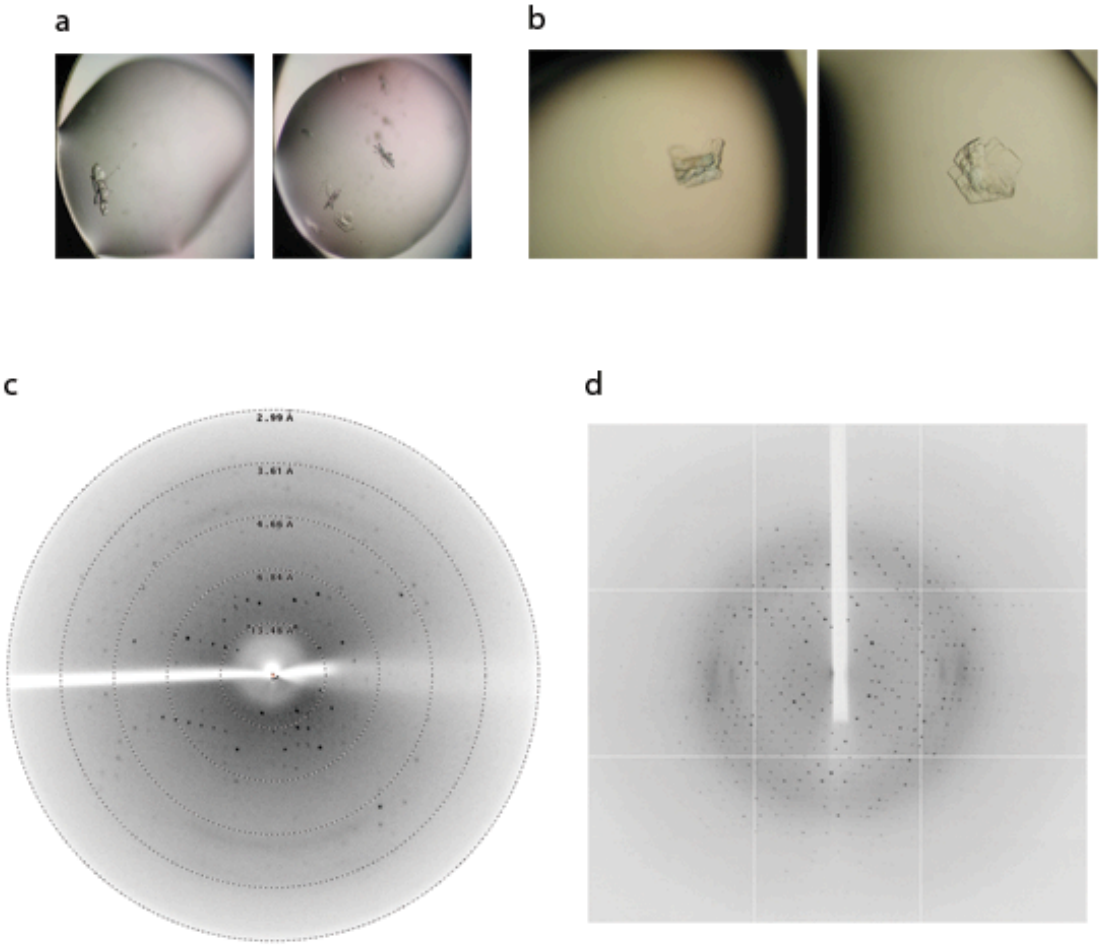


Figure 4-3: Crystals and diffraction of AgRad61$^{184-561}$

Crystals obtained in (a) initial screens and (b) after optimization. Diffraction images obtained with (c) in-house and (d) synchrotron X-ray sources. The edge of the synchrotron diffraction image shown in (d) is ~1.9 Å

4.2 AgWpl1$^{184-561}$ structure solution

4.2.1 Data collection, processing and analysis

Data for native as well as selenomethionine-labelled crystals were collected under cryo conditions at the Diamond light source IO2 beamline. Initial indexing using MOSFLM showed that the crystals belonged to the space group P1. The space group was confirmed as P1 based on the observations: (a) The spot profile of the crystals matched best with that of P1 and showed high penalty for the next space group, C2. (b) The unit-cell dimensions of these crystals (mentioned below) did not appear to match that of any other space group.

Peak and inflection data were collected for 720° with 1° oscillation whereas remote data was collected for 360° with the same oscillation. All data were then processed using MOSFLM (Leslie, 2006) in the space group P1 and scaled using SCALA (Evans, 2006). The unit cell dimensions of the native crystals were a=60.11, b=62.87, c=63.27, α=99.97°, β= 110.91°, γ=99.23°. MAD data consisting of peak, inflection and remote datasets could be collected from a single crystal as there were no apparent signs of deterioration of data quality (as can be seen from data collection statistics). The data showed good statistics with about 97% completeness and high multiplicity (7.8 for peak and inflection and 2.9 for remote) and uniform R-merge values. The I/σI values indicated usable data up to a resolution cut-off of about 2 Å for all datasets (Figure 4-4).

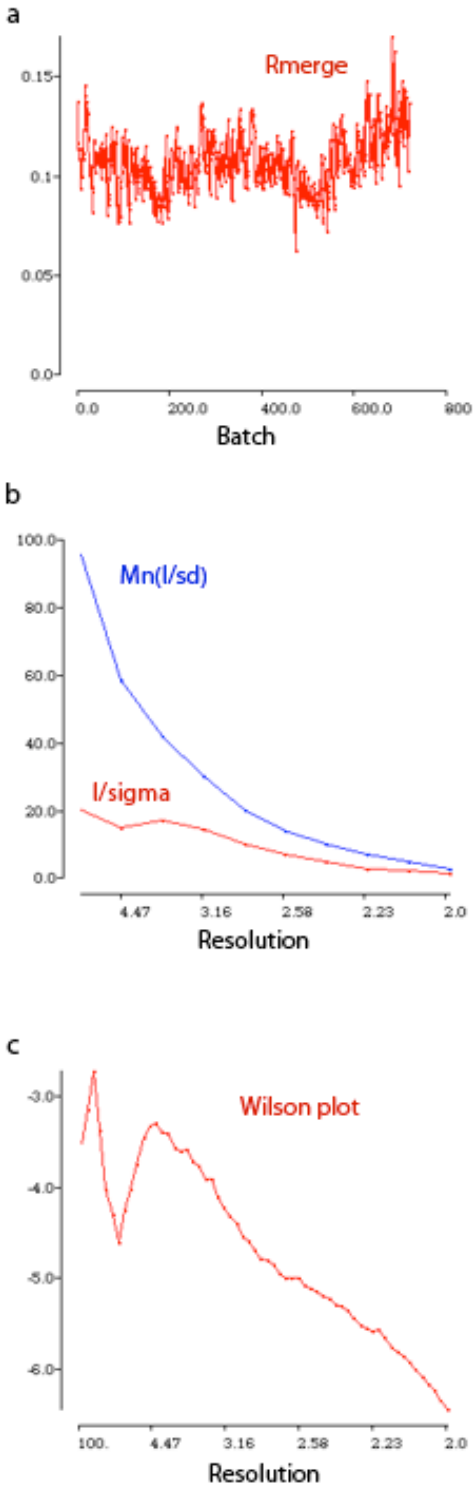


Figure 4-4: SCALA output graphs for peak dataset

Graphical representation of data collection statistics showing plots of **(a)** Rmerge vs. Batch, **(b)** I/sigma vs. Resolution, and **(c)** Wilson plot for the peak dataset. Uniform Rmerge values across batches for all datasets were observed. I/sigma values indicated usable data till 2Å resolution cut-off for the datasets collected and the Wilson plots showed a linear graph at higher resolution and slopes corresponding to reasonable B-factors (Table 4-1).

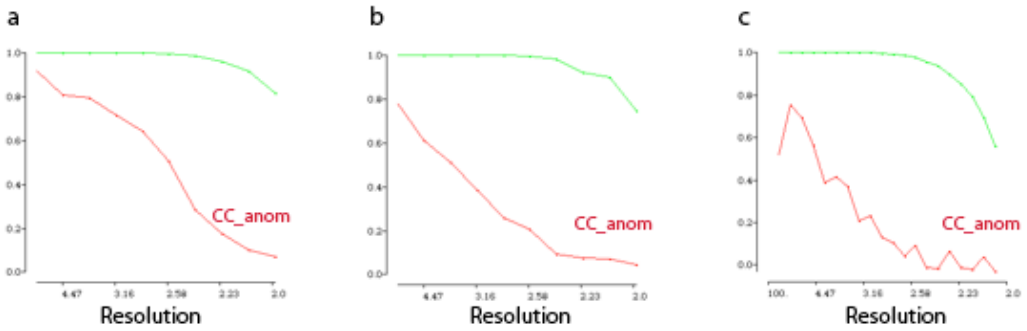


Figure 4-5: Graphical representation of anomalous signal

Signals (CC_anom) from (a) peak, (b) inflection, and (c) remote datasets produced by SCALA. A clear anomalous signal (indicated by a CC_anom value of 0.3 or more) could be observed in the peak dataset up to ~2.6Å.

Table 4-1: Data collection statistics of native and MAD datasets of AgWpl1$^{184-561}$ crystals

	Native	Peak	Inflection	Remote
Space group	P1	P1	P1	P1
Cell dimensions				
***a, b, c* (Å)**	60.11, 62.87, 63.27	60.62, 63.19, 63.5	60.7, 63.32, 63.58	60.74, 63.39, 63.63
***α, β, γ* (°)**	99.97, 110.91, 99.23	100.22, 110.65, 99.42	100.22, 110.65, 99.41	100.22, 110.65, 99.4
Wavelength (Å)	0.9720	0.97950	0.97970	0.97200
Resolution range (Å)	60.09 – 2.13	60.28-2.01	60.41-2.06	60.48-2.09
Total reflections	172038	416890	398971	141132
Unique reflections	44611	53156	50839	48714
Multiplicity	3.9 (3.8)	7.8 (7.7)	7.8 (7.9)	2.9 (2.9)
Completeness (%)	96.70 (94.20)	96.8 (93.6)	97.0 (95.9)	97.0 (96.0)
I/σ(I)	12.5 (2.10)	22.0 (2.4)	24.3 (2.9)	16.1 (2.0)
Wilson B-factor	41.06	30.159	31.825	33.415
R-sym	0.058 (0.614)	0.065 (0.726)	0.060 (0.694)	0.044 (0.485)

(figures given in parentheses refer to data in highest resolution shell)

Analysis of Matthews coefficient (2.53 for 2 molecues) for unit cell contents suggested two molecules per asymmetric unit (ASU) with 51.3% solvent content. Presence of two molecules in the ASU suggested the presence of non-crystallographic symmetry (NCS) and the data were checked for the presence of self-rotation peaks using MOLREP in CCP4 (1994). A self-rotation peak was observed at Chi=180° stereo-projection. These analyses confirmed the presence of a NCS axis with two molecules in the ASU (Figure 4-6).

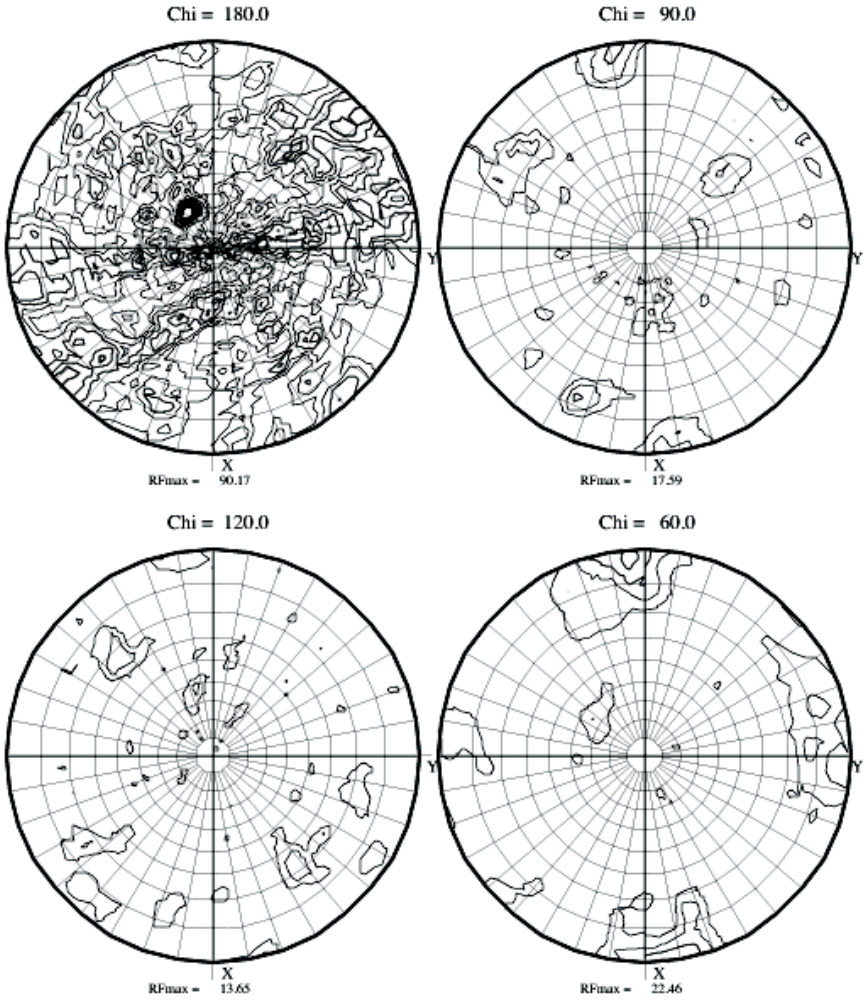


Figure 4-6: Self-rotation function analysis with MOLREP

4.2.2 MAD phasing

The anomalous data was used for structure solution using the program AutoSHARP. It is an automated structure solution suite, which combines heavy atom detection by the program SHELX (Sheldrick, 2008) and likelihood-based heavy-atom refinement by SHARP (Bricogne et al., 2003).

Processed peak, inflection and remote datasets in scalepack format were used as input for AutoSHARP. After carrying out an initial check of the supplied information, the program analysed the unit cell contents and based on the Matthews coefficient, reported two molecules in the ASU with 51.7% solvent content. Further analysis showed the presence of an NCS axis and a self-rotation peak was observed at κ=180° section (Figure 4-7).

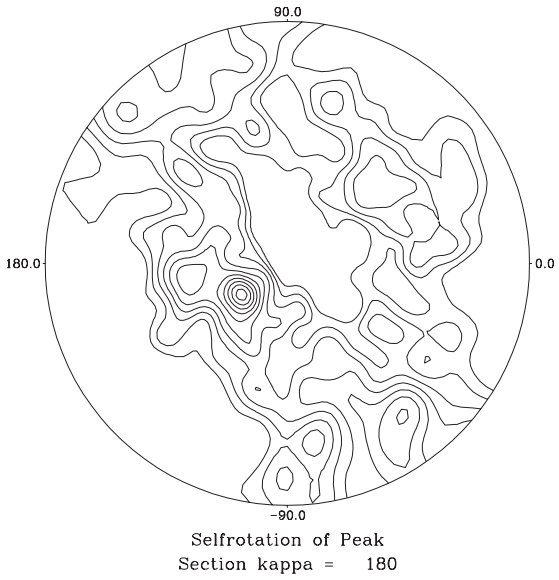


Figure 4-7: Self-rotation peak detected in AutoSHARP

The next step carried out by AutoSHARP was heavy atom detection using the program SHELX, done in a stepwise manner using the components SHELXC and SHELXD. SHELXC carried out an analysis of the data and suggested a resolution cut-off of 46.6 - 2.6Å for the anomalous signal as the correlation falls to about 30% beyond 2.6Å (consistent with that reported by SCALA; Figure 4-5). It then prepared reflection files for feeding into SHELXD based on the following 3 scenarios: (a) MAD with all three

datasets, (b) SAD with the peak dataset, and (c) MAD with first two datasets, *i. e.*, peak and inflection. SHELXD then performed a heavy atom search for each case in essentially the same order described above. The search is in a decreasing order of probability of a scenario to provide the cleanest and strongest anomalous signal. The best signal that SHELXD found was from SAD using the peak dataset, the correlation for which was 0.282. The program could detect five HA sites using the peak dataset, with a clear drop of occupancy between the fifth and the sixth sites, which were 0.85 and 0.62 respectively (Figure 4-8 and Table 4-2). The sites were further validated based on the crossword table and the scatter plots. The scatter plots show two distinct clusters, demonstrating a clear segregation of correct and wrong substructure solutions. A good correlation existed among the correct solutions, which was 45.82% against 28.66% for the weak/wrong solutions.

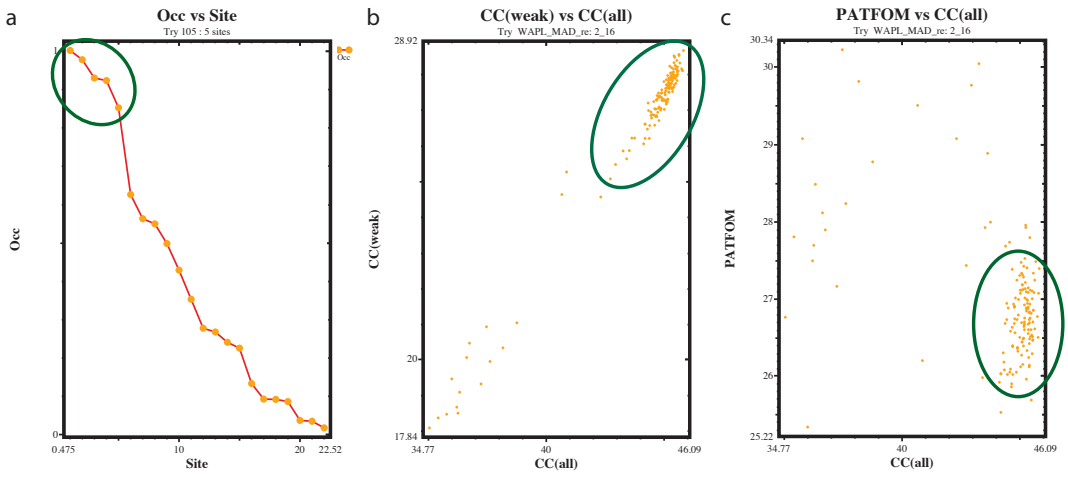


Figure 4-8:SHELXD graphical outputs

(a) Plot showing the higher occupancies of 5 (circled) out of a total 22 HA sites with a drastic decrease in occupancy of the sixth site (c) and (d) scatter plots showing a distinct cluster of correct solutions (circled) well separated from the wrong ones.

Table 4-2: Heavy atom sites detected by SHELXD

Peak	Height (rms)	X	Y	Z	X	Y	Z
		(fractional)			(orthogonal)		
1	100.00	0.0023	0.0149	0.9971	-22.39	-14.19	57.19
2	97.60	-0.3617	0.0697	1.1177	-47.70	-12.61	64.11
3	92.80	-0.4605	-0.0605	1.0191	-50.13	-19.22	58.45
4	92.20	-0.2937	0.1583	1.3757	-50.28	-11.00	78.91
5	85.10	0.0902	-0.0357	0.9246	-14.92	-16.25	53.03

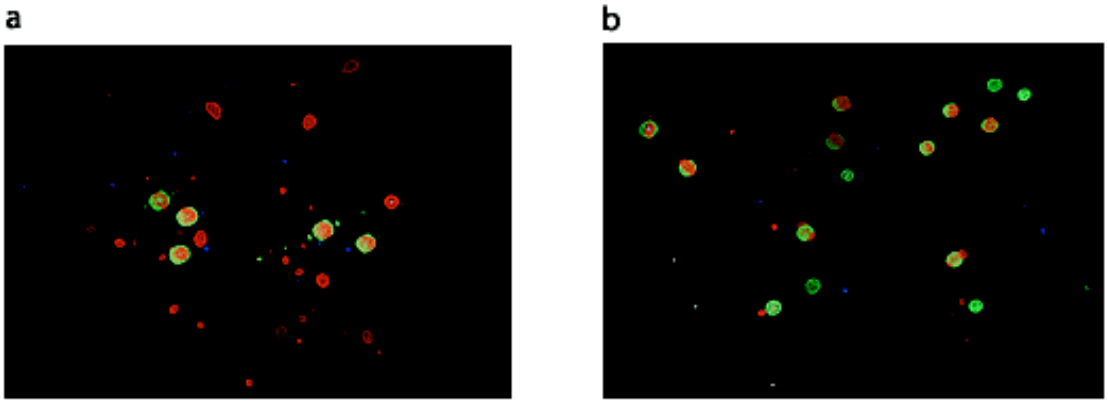


Figure 4-9: Heavy atom substructure
Residual maps, contoured at 5σ, showing heavy atom positions after **(a)** initial and **(b)** final rounds of SHARP. Green densities depict heavy atom positions already correctly assigned while red shows positions detected by SHARP for placing more potential heavy atoms.

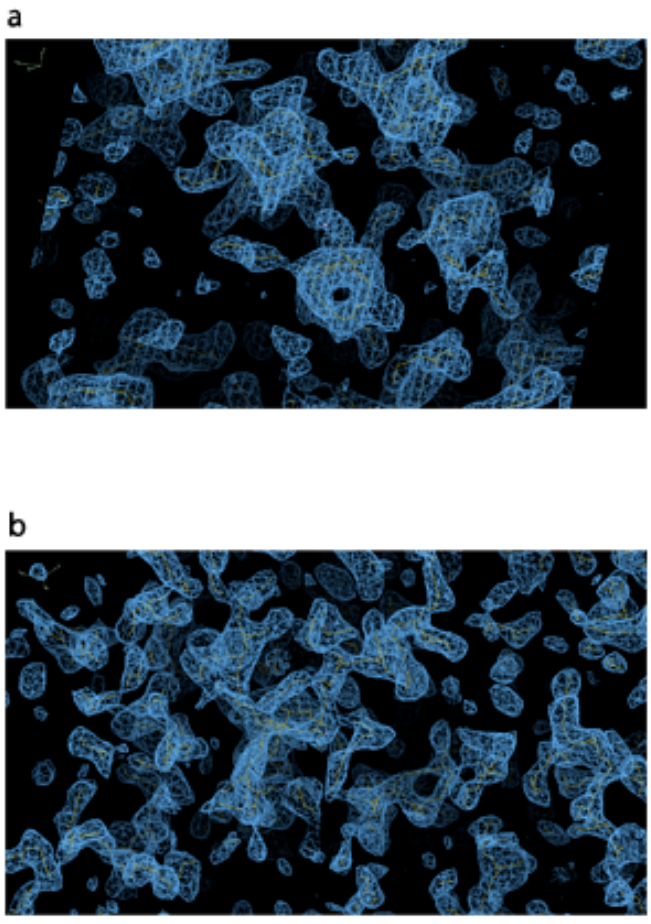


Figure 4-10: Maps generated after final SHARP run
Electron density maps obtained with heavy atom phases in (a) correct and (b) incorrect hands with clear densities for helices being observed in the former

The heavy atom coordinates from SHELXD were then used by SHARP for an initial round of refinement. After this initial SHARP run, more potential heavy atom sites could be located, as confirmed from the extra blobs of densities present in the residual maps of all the three datasets (peak, inflection and remote). Subsequently, the program could locate all 16 possible sites (8 for each monomer in the ASU) after four iterative rounds of refinement and addition/removal of the sites (Figure 4-9). An overall figure of merit (FOM) of 0.406 and phasing power of 1.109 for the Peak dataset were reported after the final round of heavy atom refinement. The phasing power fell below 1 at 2.67Å for the peak dataset, which was in good agreement with the suggested resolution cut-off by SHELXC and SCALA.

4.2.3 Model building and refinement

The density modification was carried out using the CCP4 program PARROT followed by an automated chain tracing by BUCCANEER (Cowtan, 2006). The number of amino acid residues traced by Buccaneer was 700 out of a total 756 in the ASU. After chain tracing and five rounds of refinement within Buccaneer, the R-factor was 27.98%. Further model building and refinements were done in COOT (Emsley et al., 2010) and REFMAC5 (Murshudov et al., 1997) respectively. The structure could be finally refined to an R-factor of 18.2% and R-free of 23.4%. The residues 544-548 (Chain A) and 189-194, 214-221, 311-312 (Chain B) could not be built owing to poor density in the respective regions and are missing from the final model. The model was validated using Ramachandran plot and the Molprobity tool (Lovell et al., 2003). Ramachandran plot showed 99.3% and 0.7% residues in the favoured and disallowed regions respectively.

Table 4-3: Refinement statistics

R-factor	0.182
R-free	0.234
Number of atoms	5916
Protein residues	725
Water molecules	137
RMS bonds (Å)	0.016
RMS angles (degrees)	1.81
Ramachandran favoured (%)	98.24
Ramachandran outliers (%)	0.7
Model B-factor	47.1
Molprobity clash-score	10.11

4.3 Analysis of the structure and design of mutants

4.3.1 Overall structure of the WAPL domain

The WAPL domain forms an elongated structure with approximate dimensions 95x35x35 Å made up solely of helical repeats (Figure 4-11). The molecule assumes a bent or curved morphology with a curvature of approximately 140° at the centre of the concave surface. In *A. gossypii*, the domain is made of 20 helices, 19 of which participate in the formation of 9 helical repeats. Based on overall appearance, the structure can be divided into three sub-domains – the N- and C-lobes and an elongated central region connecting the two, made up of 3 helical repeats each (Figure 4-12).

A search for structural homologues using the DALI server (Holm and Rosenstrom, 2010) showed the nearest WAPL domain homologues to be HEAT repeat proteins like β-catenin and protein phosphatase 2A which structurally align best to the middle sub-domain of WAPL (Figure 4-13). HEAT repeats are 37-47 residue long hairpin-like motifs made up of two antiparallel α-helices connected by a loop. Generally, 3-30 such repeats may be present in the proteins belonging to this family, which mostly form the protein-protein interaction modules present in many transport, signalling and scaffolding proteins. The classical examples of HEAT repeat proteins are **h**untingtin, elongation factor 3 (**E**F3), regulatory domain of protein phosphatase 2A (PP2**A**) and the PI3-kinase **T**OR1, based on which the family derives its name.

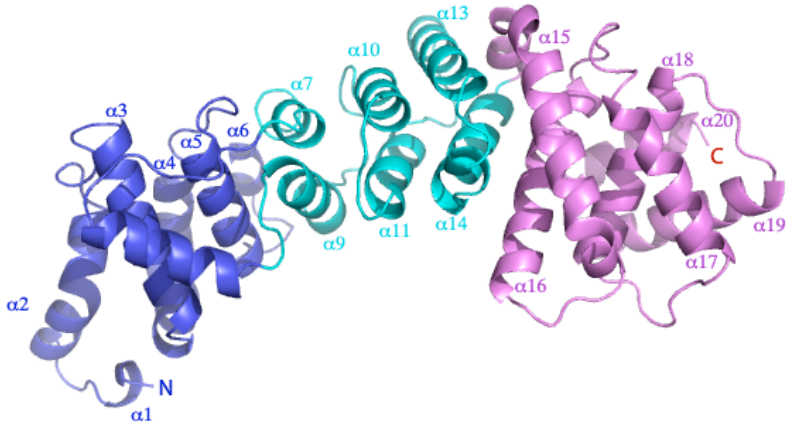


Figure 4-11: Structure of the WAPL domain

Overall structure of the WAPL domain showing the arrangement of helices labelled α1 to α20. The sub-domains are shown in different colours with the N- and C-lobes in dark blue and purple respectively and the middle region in cyan.

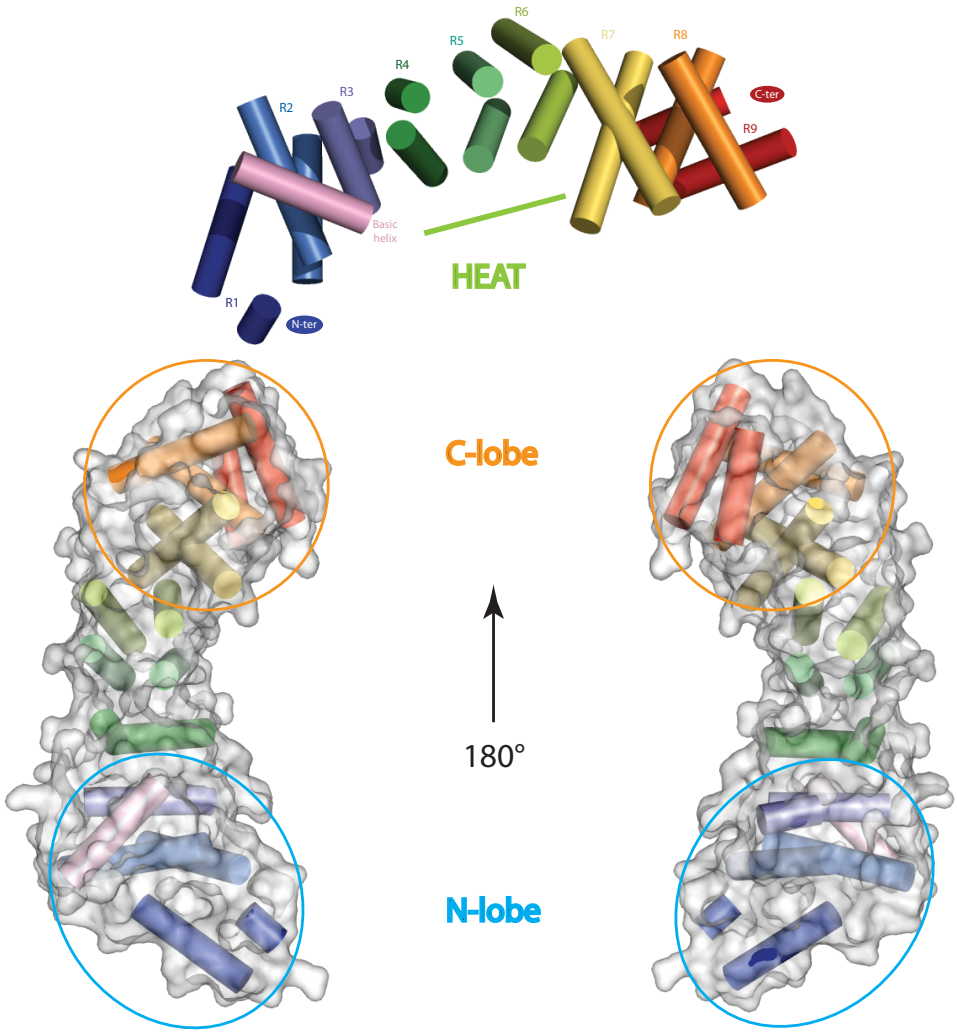


Figure 4-12: Arrangement of WAPL helical repeats

The nine helical repeats of WAPL domain arranged from N- to C-terminal are represented as cylinders and labelled as R1 to R9 (top). A pair of helices comprising a repeat has been shown in the same colour. Also shown in pink is the basic helix lying on the surface of the N-ter sub-domain. The N- and C-lobes and their surfaces are marked in blue and orange boundaries respectively (bottom).

The individual motifs comprising the N- and C-lobes of the WAPL domain roughly resemble the HEAT repeats, but are arranged as bundles of helices imparting a globular appearance to the sub-domains. This deviates from the arrangement of the motifs in canonical HEAT proteins, in which the individual repeats have a more regular arrangement resulting in an elongated morphology of the corresponding proteins. The repeats in the elongated sub-domain of WAPL assume an arrangement much more

similar to that of HEAT–repeat proteins. It consists of 3 helical repeats with the first and second, as one moves from N- to C-terminal, being made of much shorter helices but longer and irregular loops than usually found in these repeats. The first repeat is especially unusual as the connecting residues between the two helices that constitute the motif form an additional helix (residues 312-324) with extended loops on either side connecting it to the two main helices. The long loops cause a protrusion of the helix and make it traverse all the way to the surface of the N-lobe. The helix and its flanking loops form a lysine and arginine - rich stretch imparting a localized positive charge to the N-lobe.

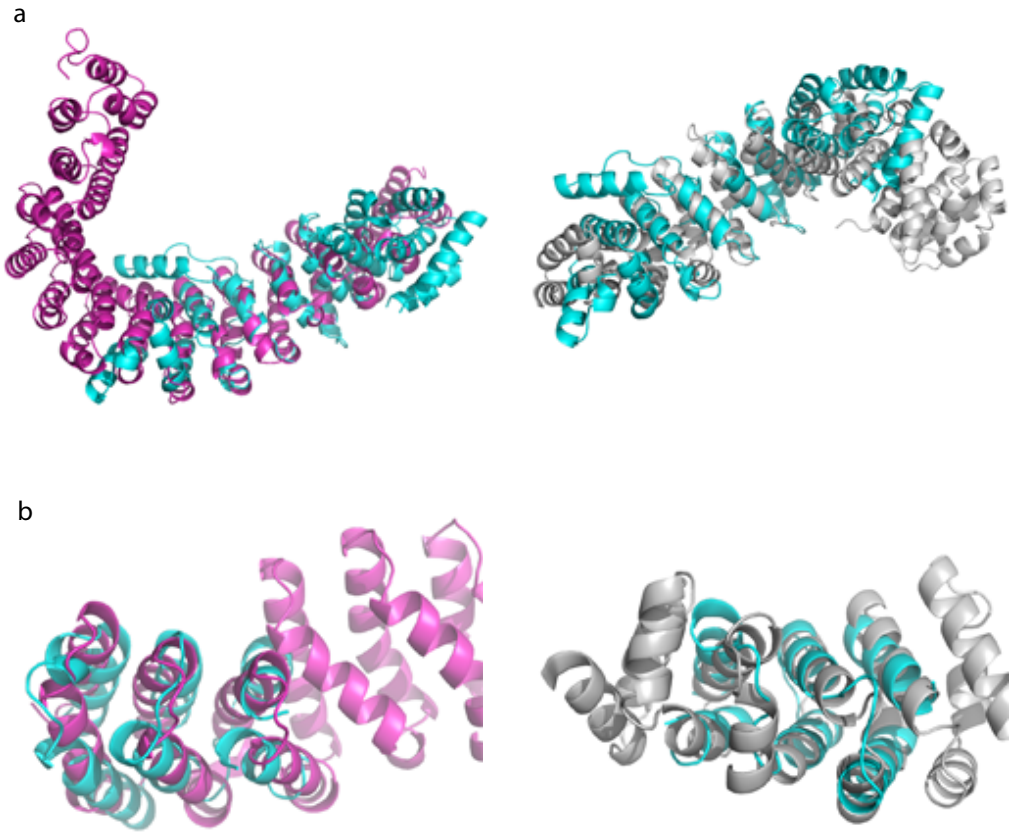


Figure 4-13: Structural alignments of WAPL domain with canonical heat repeat proteins

Alignments of PP2A [PDB: 1B3U] and β-catenin [PDB: 3BCT] with (a) full-length and (b) only the middle, residues 290-419, of WAPL domain. WAPL is shown in cyan, PP2A in purple and β-catenin in grey.

4.3.2 Surface charge distribution and conserved motifs of the WAPL domain

The WAPL domain shows scattered acidic, basic and hydrophobic patches. The N-terminal surface of the domain has a distinct acidic patch while the surface formed by the basic helix (discussed above) imparts an extended basic surface to this region. The middle and the C-terminal of the protein show a prevalence of negatively charged and hydrophobic regions (Figure 4-14a).

Although the WAPL domain is conserved throughout eukaryotes, the fungal orthologues, including that of *Ashbya* and budding yeast, show a low sequence similarity to the WAPL of higher eukaryotes, like Drosophila and human. Due to this reason, a sequence alignment of different fungal WAPL orthologues was carried out and the conserved stretches were mapped onto the structure of AgWpl1$^{184-561}$. Based on the alignment, five conserved stretches of the WAPL domain were identified and named Motifs I to V (Figure 4-14b and c). All these motifs could be mapped to surface exposed areas on the molecule. Motif I comprises of a conserved acidic stretch in the extreme N-terminal end of the domain while Motif II is made up of the basic helix and the adjoining loops. Motifs I and II account for the prominent acidic and basic patches, respectively, seen on the N-terminal surface of the WAPL domain. Motifs III, IV and V are present toward the C-terminal of the protein and are mostly made of hydrophobic stretches. These three motifs collectively form a hydrophobic patch on the concave surface of the WAPL C-terminal and together they also contribute toward the formation of a hydrophobic pocket.

4.3.3 Design of WAPL domain mutants

WAPL mutants for further biochemical analyses were designed based on the conserved structural features as a guide. Most of the mutations were designed for *S. cerevisiae* WAPL domain, as further biochemical experiments were mostly performed using this orthologue. The mutations and the corresponding motifs they are located in have been listed in Table 4-4.

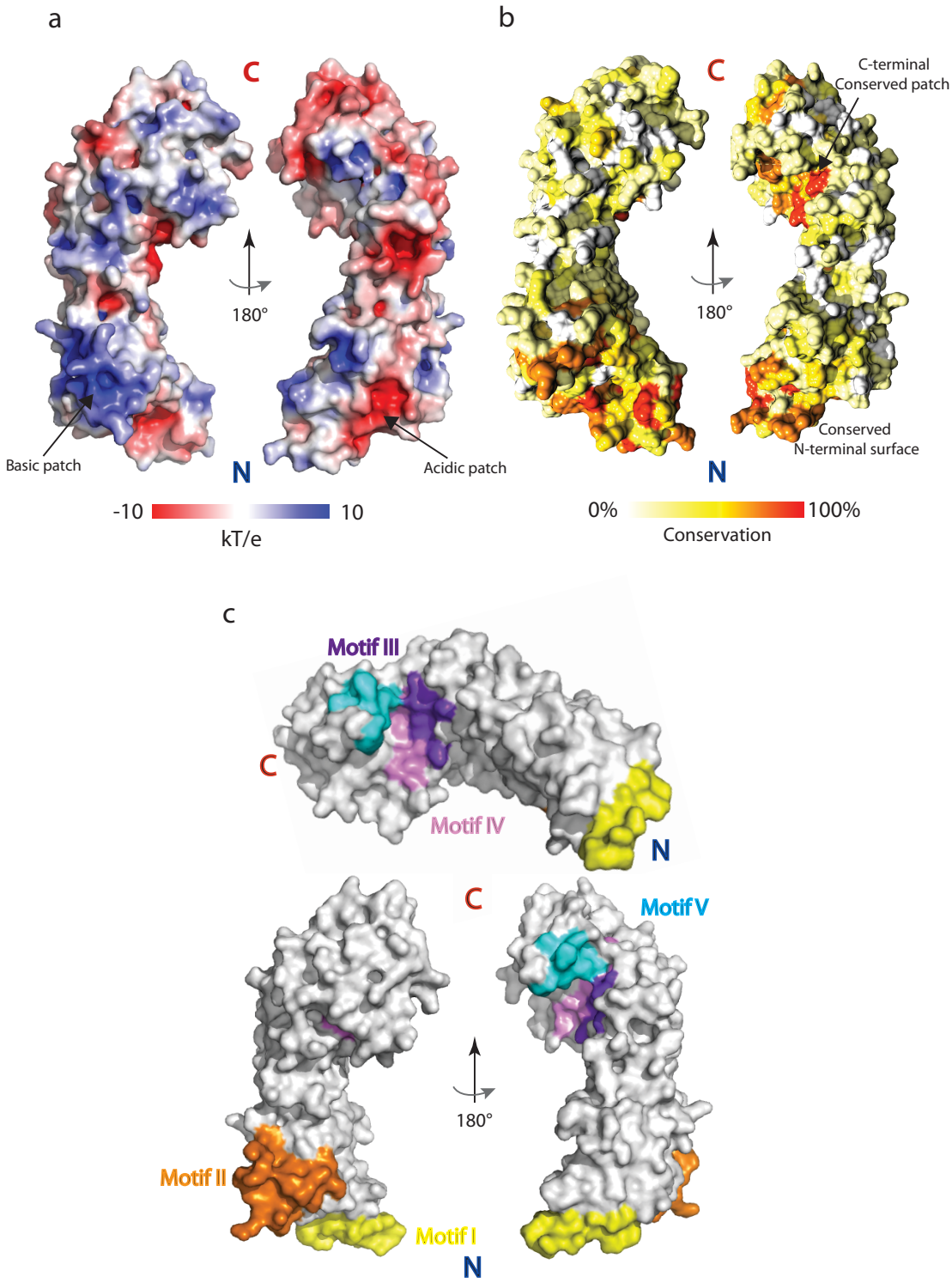


Figure 4-14: Charge distribution and conservation on the WAPL domain surface

Surface charge distribution of WAPL domain (a) and conservation profile calculated using fungal WAPL sequences and conserved motifs I to V mapped onto the surface of AgWpl1$^{184-561}$ structure (b) & (c)

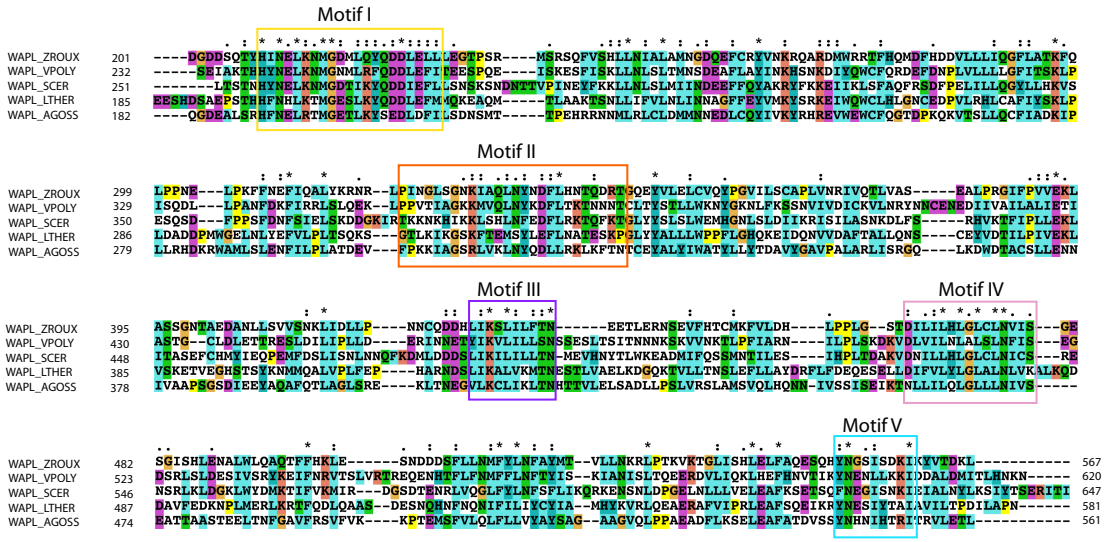


Figure 4-15: Multiple sequence alignment of fungal WAPL domains

Alignment of WAPL domain sequences from the fungi Z. rouxii, V. polyspora, S. cerevisiae, L. thermotolerans, and A. gossypii. The positions of the conserved motifs have been indicated with boxes (color coded as in Figure 4-14c)

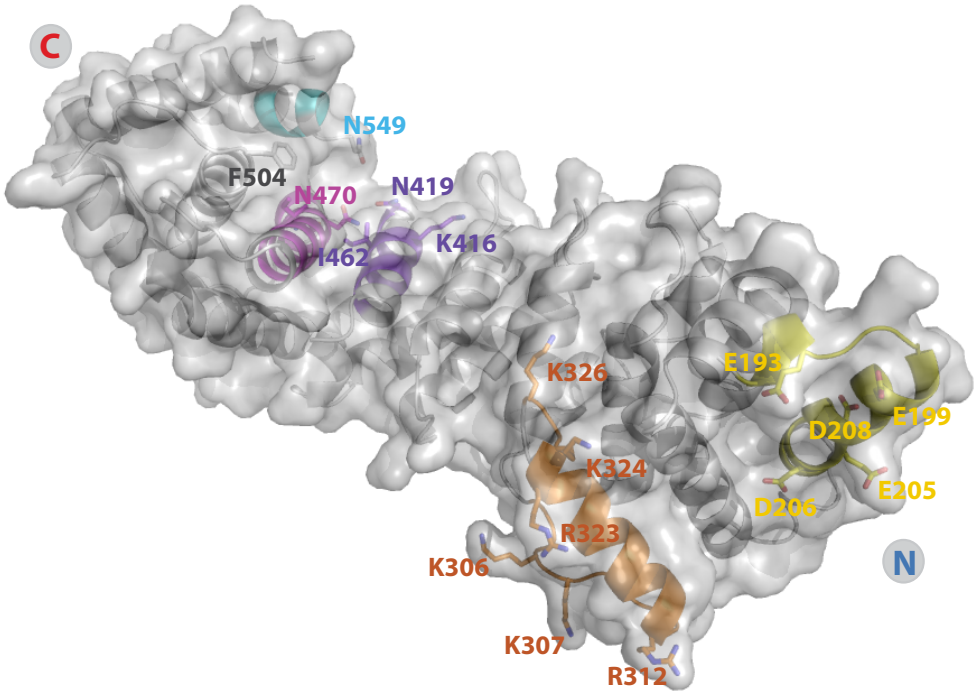


Figure 4-16: Surface-rendered view of the WAPL domain

Side chain conformation of some of the conserved and surface-exposed residues in the identified motifs

Table 4-4: Conserved motifs and mutations

Motifs	Residue range		Mutations (in *S. cerevisiae*)
	A. gossypii	*S. cerevisiae*	
I	190-210	256-276	D271, 272A
II	304-325	375-396	K376, 377, 382 and 383Q
III	410-419	484-493	K486N/R and N493A/E
IV	459-473	529-543	G536W and N540A
V	546-554	621-629	N622A/E

Motif I mutations:

The residues D271 and D272 of ScWpl1 Motif I are totally conserved among fungal orthologues except that in *A. gossypii* the first aspartate is replaced by a glutamate (E205/D206). These residues in ScWpl1$^{259-647}$ were mutated to alanine to verify the effect of this localized and conserved negatively charged surface.

Motif II mutations:

This motif has a series of surface exposed lysines and arginines in all the orthologues but the positions of these residues within the motif are not exactly conserved. So, the surface exposed lysines in ScWpl1$^{259-647}$ Motif II, which included K376, K377, K382 and K383, were changed to glutamines.

Motifs III, IV and V mutations:

These three motifs together make up the conserved hydrophobic inner surface on the WAPL domain C-terminal, as discussed earlier. The mutations included K486N/R and N493A of Motif III, G536W and N540A of Motif IV and N622A of Motif V.

Chapter 5. Functional analyses of the WAPL domain

The helical repeat structure of WAPL and its structural homology to HEAT repeat proteins suggest that this domain might be involved in protein-protein interactions. Identification and characterization of such interactions within the context of sister chromatid cohesion might provide deeper insights into the poorly understood role of Wpl1 as a cohesin regulator. So, experiments were carried out to identify as well as validate interactions mediated by this conserved domain. The present chapter describes the rationale and details of the experiments undertaken toward this goal. Furthermore, the functional characterization and verification of the interactions both *in vitro* and *in vivo* have been discussed.

5.1 Peptide array

Peptide array experiments were carried out as screens to search for WAPL domain interacting partners within the cohesin subunits. The proteins that were screened included *A. gossypii* orthologues of the Smc1 (residues 1-190 and 1045-1222) and Smc3 (residues 1-190 and 1045-1231) ATPase head domains and Scc1. The Pds5 and Scc3 subunits were left out as they have been shown to interact with the N-terminal of Wpl1 in previous studies. The above-mentioned proteins were arrayed as 21-mer peptides with a shift of 3 residues (18 residue overlap) between successive peptide spots. The arrays were probed with AgWpl1$^{184-561}$ protein in order to identify interactions mediated solely by the WAPL domain. A concentration of 50 nM AgWpl1$^{184-561}$ was used to probe the arrays and an α-AgWpl1$^{184-561}$ antibody was used to detect any binding. Simultaneous experiments in the absence of AgWpl1$^{184-561}$ served as negative controls and served to blank out any non-specific signal arising from the antibodies.

After comparisons with the respective blanks, potential hits could be observed in the Smc3 ATPase head and Scc1 arrays. There were 3 positive hits each in the Smc3 ATPase head and Scc1 arrays but only the Smc3 hits were further characterized. The positions of these peptides in Smc3 and Scc1 primary sequence have been depicted in Figure 5-1(b).

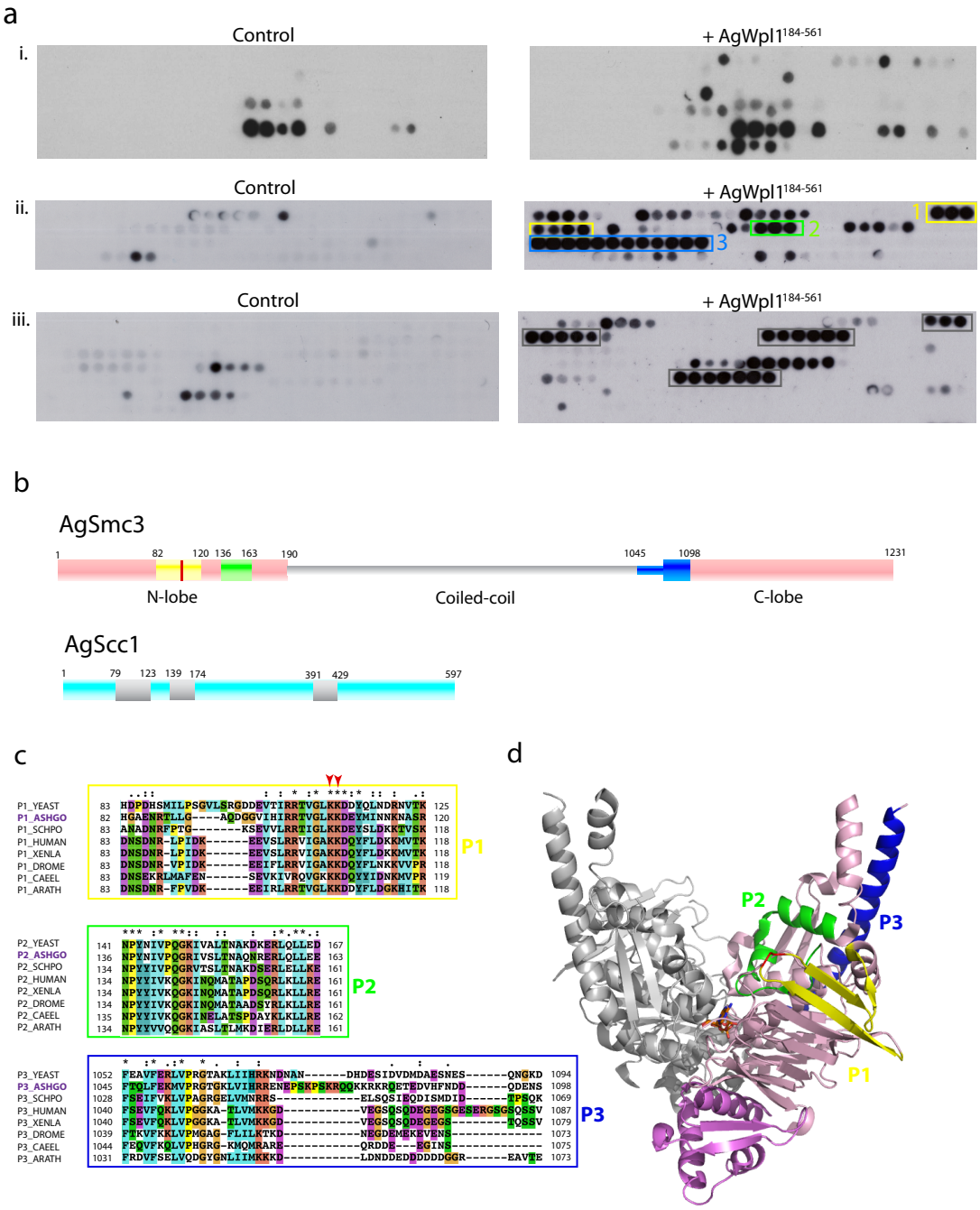


Figure 5-1: Peptide array hits

(a) Peptide array scans of AgSmc1 head (i), AgSmc3 head (ii) and AgScc1(iii). The controls for each are shown in the left panel while in the right are the arrays after probing with AgWpl1$^{184-561}$. The positive hits in the AgSmc3 head are boxed and numbered (AgScc3 hits only boxed). **(b)** The positions of the positive hits of AgSmc3 heads and AgScc1 shown in the primary structure of the respective full-length proteins. **(c)** Multiple sequence alignments of the Smc3 positive hits P1, P2 and P3 using the respective stretches of the Smc3 orthologues of yeast, *A. gossypii*, *S. pombe*, human, *X. laevis*, *D. melanogaster*, *C. elegans*, and *A. thaliana*. **(d)** Mapping of the P1, P2 and P3 stretches on the ScSmc1 ATPase–Scc1 (C-ter) complex dimer structure (PDB code: 1W1W; an SCC1 monomer is depicted in purple). The positions of the hits are depicted in only one monomer. An uniform colour scheme for the hits P1 (yellow), P2 (green) and P3 (blue) has been used.

The AgSmc3 positive hits, termed P1 (residues 82-120), P2 (residues 136-164) and P3 (residues 1045-1098), were first mapped onto the previously elucidated crystal structure of budding yeast Smc1 ATPase head (PDB code: 1W1W; (Haering et al., 2004)), using the homology between Smc1 and Smc3 sequences. Interestingly, all the hits could be mapped to surface-exposed regions in the Smc1 structure with P1 and P2 being located in the N-terminal lobe and the P3 sequence toward the C-terminal half of Smc1 [Figure 5-1(b)]. The P1 stretch was found to be of particular interest as it contained the pair of lysines targeted by Eco1 for acetylation. In the Smc1 structure, it maps to a region consisting of two surface-exposed β-strands with the Eco1 target lysines projecting out of the connecting loop in between the strands. The position of peptide P2 also appeared significant as the corresponding region in the Smc1 head forms a pair of helices directly above the conserved Q-loop, which is known to bind to ATP. The sequence P3 corresponds to the base of the coiled coil preceding the C-terminal lobe in the Smc1 head structure [Figure 5-1(d)].

A multiple sequence alignment of diverse Smc3 orthologues showed a good conservation of the P1, P2 and P3 stretches [Figure 5-1(c)]. Based on the alignment, P1 showed a short variable portion on its N-terminal while P3 had a poorly conserved region toward its C-terminal. The alignment was used to design peptides for further binding studies described in the following sections.

5.2 Fluorescence polarization

Fluorescence polarization (FP) assay was carried out in order to check the validity of the hits obtained from the peptide array experiment. In order to use optimal lengths of the positive hits for this assay, the most conserved stretches of P1, P2 and P3, were used. So, the variable regions of P1 and P3 (as mentioned before) were omitted from the peptides that were synthesized. The optimal peptide sequences used are shown in Figure 5-2 and were named WIS1, WIS2 and WIS3 (***W***APL ***i***nteracting ***s***equence).

The WIS peptides were synthesized as fluorescein-tagged peptides and their binding to the AgWpl1$^{184\text{-}561}$ protein was tested. The assay was performed in 20 μL volumes in which peptides were added to a fixed final concentration of 50 nM. The protein concentration was varied such that a saturation of binding could be obtained thus allowing the calculation of the dissociation constant (K_d). The experiment was performed in triplicates for each concentration of the protein that was used.

5.2.1 Assay optimization

An optimization of the assay buffer was first carried out in order to obtain dependable and reproducible binding curves. Experiments carried out in buffers containing higher than 20 mM salt were difficult to interpret, as in such cases, saturation of binding could not be achieved even with the highest achievable concentrations of AgWpl1$^{184\text{-}561}$ (~15 mg/mL). Based on these initial trials, the final buffer composition used for this assay could be optimised, which was 20mM Tris pH 7.5, 20 mM NaCl and 0.5 mM TCEP. The peptides and protein used in the assay were dialysed into the above buffer before performing the assays. Following buffer optimization, the range of protein concentration to be used was tested and a range of 0 to 300 μM was found suitable given the buffer composition and the peptides used for the assay.

a

WIS1: 101 - RRTVGLKKDETM - 112

WIS2: 136 - NPYNIVPQGRIVSLTNAQNRERLQLLEE - 163

WIS3: 1045 - FTQLFEKMVPRGTGKLVIHRRE - 1066

b

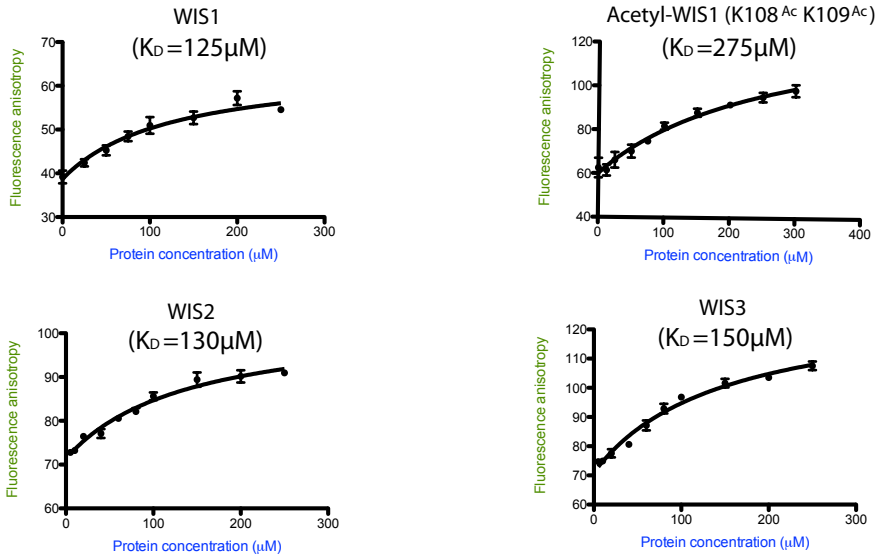


Figure 5-2: WAPL domain binding with WIS peptides

(a) The optimized sequences used for the FP assays, named WIS1, 2 and 3. The start and end residue numbers are indicated. **(b)** FP assay based dissociation curves obtained for the respective wild-type peptides and the acetylated WIS1 peptide The peptides and their K_d are indicated on the respective curves.

5.2.2 Binding affinities of WAPL domain with WIS peptides

All the three peptides – WIS1, 2 and 3 were found to bind with micromolar affinities showing dissociation constants 125, 130 and 150 μM respectively [Figure 5-2(b)]. The results showed that the potential hits obtained from the peptide arrays were valid and are probable sites of interaction between the respective proteins.

The WIS1 sequence contains the pair of lysines (K105/K106 in AgSmc3 and K112/K113 in ScSmc3) that act as Eco1 acetylation targets. In order to test whether the acetylation of these lysines affect the binding of the WIS1 peptide to $AgWpl1^{184-561}$, a variant peptide carrying ε-acetyl lysines replacing the K105/K106 was synthesized and its binding checked. The dissociation constant in this case was found to be 275 μM, which shows the binding affinity upon acetylation is reduced by half compared to the wild-type peptide.

The difference in binding affinities of the WAPL domain with wild type and acetylated forms of the WIS1 sequence is quite significant, as this would explain how Eco1-mediated acetylation of Smc3 *in vivo* overcomes the destabilizing effect of Wpl1 on cohesin. Nevertheless, the results needed to be further verified in the context of intact proteins and has been discussed in the next section.

5.3 WAPL–WIS2 complex structure

As a further step towards the characterization of Wpl1-Smc3 interaction, it was important to identify regions of the WAPL domain that bind Smc3. In order to answer this, co-crystallization of the AgWpl1184-561 and the WIS peptides were carried out. These crystal trials were set-up along with Dr. Xiao Wen-Hu.

5.3.1 Crystallisation

The protein used for the crystal trials was dialysed into the buffer 20mM Tris pH 7.5, 20mM NaCl and 1mM DTT with the peptide stocks prepared in the same buffer as well. WIS1, WIS2 and WIS3 peptides at 5X and 10X molar excess were added to the protein at 5mg/mL (~116 μM) and were incubated at 4°C for 30 minutes before setting up the crystallization screens. The screens were set-up in 96-well sitting drop plates with 400 nL drops made up of equal volumes of protein and the well-solution. Diffraction quality crystals of AgWpl1$^{184-561}$ could be obtained only in the presence of the WIS2 peptide. The best quality crystals, with plate-like morphology, were obtained in the condition 0.2M ammonium sulphate, 0.1M Tris pH8.5 and 25% (w/v) PEG3350. These crystals were mounted on the X-ray beam after flash-freezing in the mother liquor supplemented with 30% glycerol and diffracted to 2Å in-house under these conditions.

5.3.2 Data collection and processing

Diffraction data for the AgWpl1$^{184-561}$-Smc3^{WIS2} complex crystal was collected using the in-house X-ray source. Initial indexing in MOSFLM showed the space group to be C2. Data was collected for 187º of crystal rotation with 8 minutes exposure and 1º oscillation per frame. However, the program Pointless suggested an I2 setting (owing to a relatively smaller β angle) of the space group C2 and therefore the space group was reindexed to I2. The data was integrated and scaled using XDS (Kabsch, 2010) and SCALA respectively. After data processing, the unit cell dimensions were found to be a=97.4, b=35.2, c=117.38 and α=90°, β=101.9° γ=90°. The data showed good overall collection statistics with uniform R_{merge} values, satisfactory completeness (98.7%) and

high multiplicity (3.6). I/sigma was found to be 3.3 for the highest resolution shell (2.01 Å).

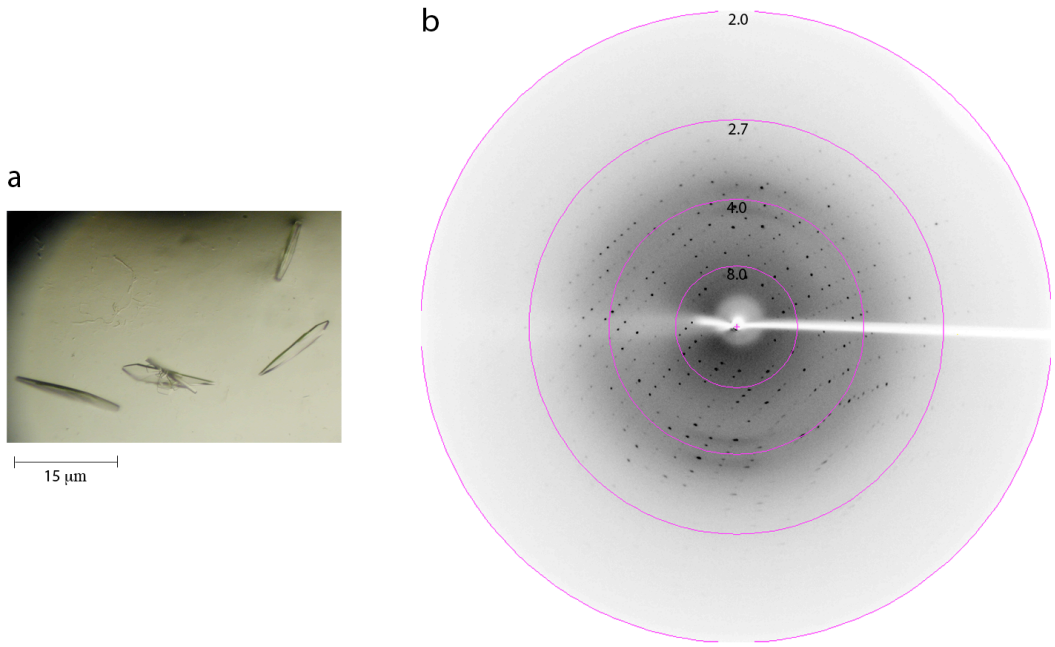


Figure 5-3: Crystals and diffraction of AgWpl1$^{184-561}$-WIS2 complex

Table 5-1: Data collection statistics of AgWpl1$^{184-561}$-WIS2 complex crystal

Space group	I2
Cell dimensions	
a, b, c (Å)	97.4, 35.2, 117.38
α, β, γ (°)	90 101.89, 90.00
Wavelength (Å)	1.54
Resolution range (Å)	82.14-2.01
Total reflections	94341
Unique reflections	26061
Multiplicity	3.6 (3.1)
Completeness (%)	98.7 (94.3)
I/σ(I)	12.3 (3.3)
Wilson B-factor	20.9
R-sym	0.075 (0.407)

(figures shown in parentheses refer to data in highest resolution shell)

5.3.3 Molecular replacement phasing

The AgWpl1$^{184-561}$-Smc3^{WIS2} complex structure was solved by molecular replacement using the program PHASER (McCoy et al., 2007). Matthews coefficient (2.34 for 1 molecule in the ASU; P(tot)=1.00) analysis showed 1 molecule of the complex in the ASU with 47.6% solvent content. So a single copy of native AgWpl1$^{184-561}$ structure was used as a search model for the molecular replacement. The intact WAPL domain, when used as the search model, did not yield a clear solution. It was reasoned that this might be due to the flexibility in the WAPL domain, which might cause a conformational change upon binding to the peptide. So, the molecule was divided into 3 parts based on the subdomains described before. The residue ranges were – A (Ensemble 1): 189-292, B (Ensemble 2): 293-432 and C (Ensemble 3): 433-561. The respective coordinates were used as discrete ensembles for phasing of the complex using the AutoMR mode of PHASER. The program reported a single solution with a log-likelihood gain (LLG) score of 1604 and translation function Z score (TFZ) of 36.2, which usually suggests a good solution.

Table 5-2: Summary of MR solution by PHASER

	Rotation function peaks			**Translation function peaks**			**LLG**	**TFZ**
	Euler1	Euler2	Euler3	X	Y	Z		
Ensemble 1	220.3	88.1	218.1	0.51	-0.05	0.59	1604	36.2
Ensemble 2	223.1	86.0	215.4	0.52	-0.01	0.57		
Ensemble 3	311.2	90.8	32.5	-0.01	-0.54	-0.07		

5.3.4 Model building and refinement

The phased molecular replacement map was used for chain tracing using Buccaneer. The program could build the whole molecule and at this stage, a model with R_{factor} and R_{free} of 23 and 27% respectively was obtained. After a round of rigid body refinement using this model and corresponding phases, a clear positive Fo-Fc density (contoured at 3σ) could be observed near to the C-terminal conserved hydrophobic pocket. Subsequently, side-chains of the residues 153-163 of WIS2 could be unambiguously fitted into this density (Figure 5-4). The model could be finally refined to an R_{factor} and

R_{free} of 19.2 and 23.8% respectively and Ramachandran plot showed 98.2% residues in the favoured regions.

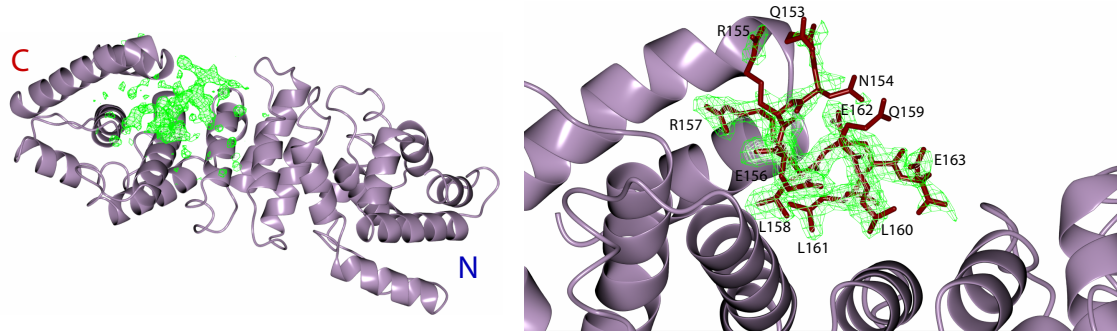


Figure 5-4: Fo-Fc density, contoured at 3σ, of WIS2 peptide bound to AgWpl1$^{184-561}$ (left) and the final 2Fo-Fc density with the fitted WIS2 sequence is also shown (right).

Table 5-3: Refinement statistics of AgWpl1$^{184-561}$-WIS2 structure

R-factor	0.192
R-free	0.238
Number of atoms	3031
Protein residues	373
Water molecules	58
RMS bonds (Å)	0.02
RMS angles (degrees)	2.02
Ramachandran favoured (%)	96.8
Ramachandran outliers (%)	1.36
Model B-factor	29.7
Molprobity clash-score	9.18

5.3.5 Analysis of AgWpl1$^{184-561}$-WIS2 complex structure and design of mutants

The WIS2 peptide was bound to the conserved C-terminal surface (formed by Motifs III, IV and V) with four of the residues (159-162) forming an ordered helical turn. The conserved leucines in the WIS2 peptide (L158, L160 and L161) undergo hydrophobic interactions with residues in the hydrophobic pocket, mainly I415 and I462 (parts of Motifs III and IV respectively). The side chains of the WIS2 leucines intercalate into the hydrophobic core of the pocket and thus appear to be the main interaction holding the peptide to the WAPL domain. Interactions outside of the pocket include stacking of the terminal amino groups of a conserved arginine R157 of WIS2 against the phenyl ring of F504 of AgWpl1$^{184-561}$. In addition, hydrogen bonds are formed by the side chain O of E162 (WIS2) with main chain amino group of N549 (AgWpl1$^{184-561}$) and main chain carbonyl group of E163 (WIS2) with side chain amino group of N419 (AgWpl1$^{184-561}$). A summary of the interactions is listed in Table 5-4.

Structure-based mutations were designed both in the WAPL domain and the WIS2 region of Smc3 in order to verify the interactions. These mutations were made in the proteins ScWpl1$^{259-647}$ and ScSmc3 ATPase head (residues 1-190 and 1040-1230) and were aimed to disrupt the hydrophobic interactions of the pocket by mutating the participating hydrophobic residues to charged ones. The Wapl domain residues L532 and L577 were changed to glutamate and arginine respectively. The L163 and L166 of Smc3 were mutated to alanines as well as aspartates while the residue D168 was mutated to alanine.

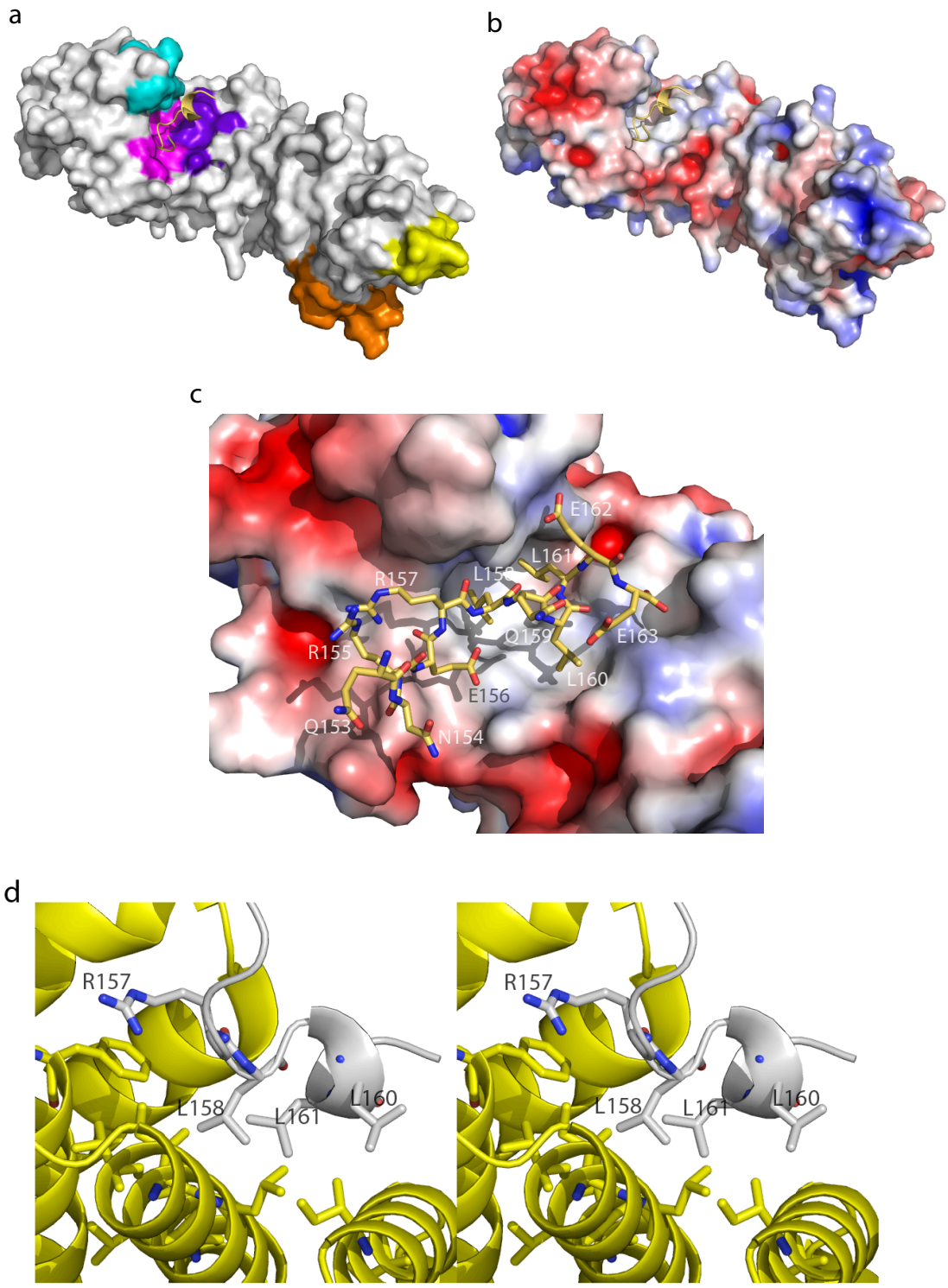


Figure 5-5: AgWpl1$^{184-561}$-WIS2 complex structure and interactions.

(a) and **(b)** The WIS2 peptide in its WAPL domain-bound state showing the position of the peptide relative to the conserved motifs and against the charged surface of the domain. **(c)** Close-up view of the hydrophobic pocket with bound WIS2 peptide showed with its side chains. **(d)** Stereo-view showing the conserved leucines of WIS2, L158, L160 and L161, stacked against the hydrophobic residues of the pocket.

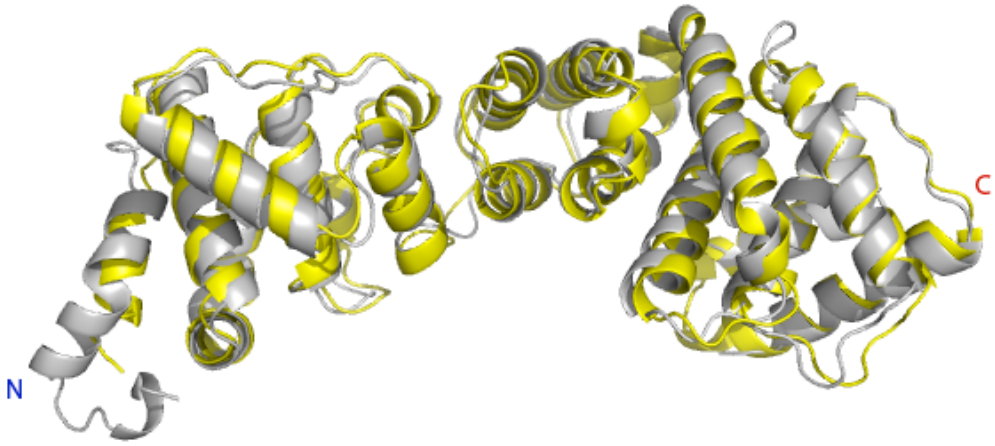


Figure 5-6: Alignment of the WAPL domain from native and complex structures

The WAPL domain from both the structures aligned with an overall RMSD of 1.384, thus showing that the domain undergoes minimal conformational changes upon binding to WIS2. WIS2 bound Wpl1 is shown in yellow whereas the native structure is shown in grey.

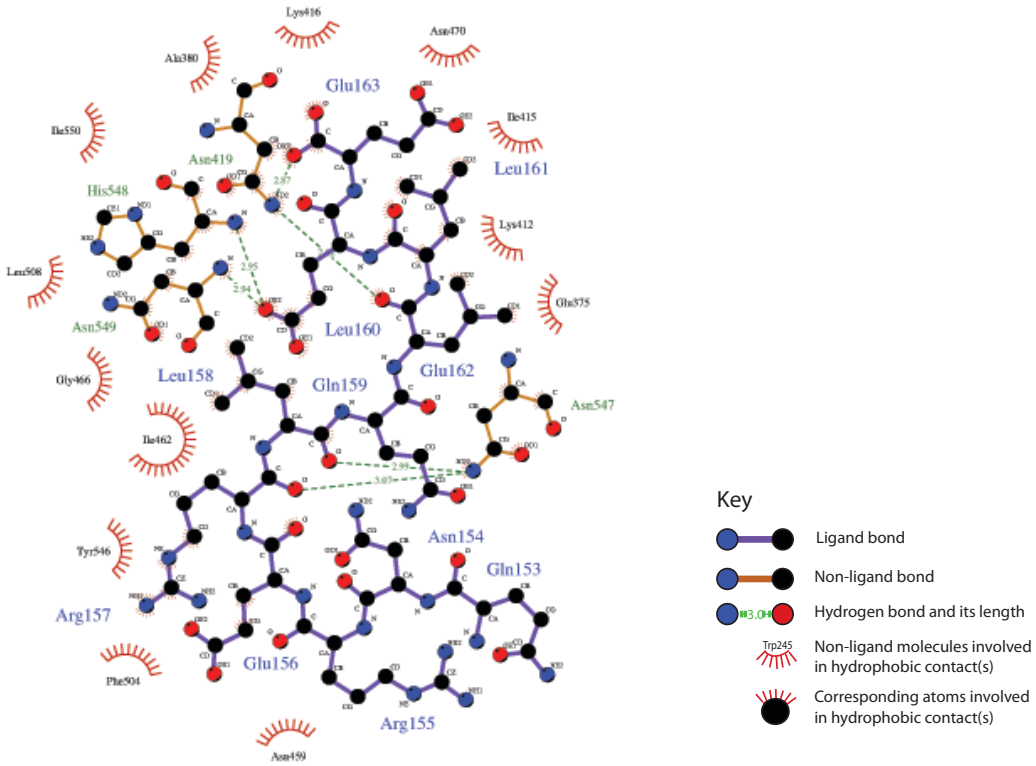


Figure 5-7: Interactions between WIS2 and the WAPL domain (using Ligplot)

Table 5-4: Summary of WAPL-Smc3^{WIS2} interactions and structure-based mutations in *S. cerevisisae* proteins

	WIS2 residues		WAPL residues		Mutants	
	A. gossypii	*S. cerevisiae*	*A. gossypii*	*S. cerevisiae*	ScSmc3	ScWPl1
Hydrophobic interactions	L158, L160, L161	L163, L165, L166	I415, I462	I489, L532	L163A/D L166A/D	L532E
Stacking interactions	R157	R159	F504	L577		L577R
Hydrogen bonds	E162 (side chain) E163 (main chain >C=O)	E167 D168	N549 (main chain $-NH_2$) N419 (side chain $-NH_2$)	G624 N493	D168A	N493E

5.4 Binding studies between the WAPL domain and Smc3 ATPase

The interaction between the WAPL domain and the Smc3 ATPase was tested both by *in vitro* and *in vivo* studies. The binding was initially verified using surface-biolayer interferometry technique in order to analyse and validate the interactions with intact and folded proteins (rather than peptides). The *in vivo* tests were done in budding yeast in order to both validate as well as verify the significance of the interaction in the cells.

Binding was tested between wild type and mutant versions of the respective proteins, allowing verification of the effects of the mutations. In addition, the relative contributions of the different binding interfaces toward the overall affinity between the proteins could also be verified by mutating key residues of different motifs in both proteins (e.g., the interfaces formed by WIS1, WIS2 and WIS3). The design of the mutations were as discussed in Sections 4.3.3 and 5.3.5.

The Smc3 constructs (wild-type and mutants) were cloned, expressed and purified by Dr. Silva Zakian and Dr. Martin Singleton carried out the *in vivo* assays.

5.4.1 Biolayer interferometry based binding studies

For these binding experiments, a truncated construct of GST-tagged ScSmc3 (hereafter called ScSmc3-ATPase) was used, which consisted of the N-lobe (residues 1-191) and the C-lobe (residues 1047-1230) of the ATPase domain together with short stretches of the coiled-coil regions and a short linker connecting the two lobes (Fig. 5-7).

The experiments were carried out by amine-coupling the WAPL domain to biosensor tips at a concentration of 12.5 µg/mL and the Smc3 head was kept in solution in concentrations ranging from 0 to 1200 nM. The results of these experiments are listed in Table 5-4. The wild type proteins – ScWpl1$^{259-647}$ and Smc3-ATPase showed dissociation constant (K_d) of 47 nM, a much tighter binding compared to the micromolar binding affinities that was observed in the FP assays. This was as

anticipated due to an extended binding interface which together contribute toward the binding affinity when proteins are used rather than peptides.

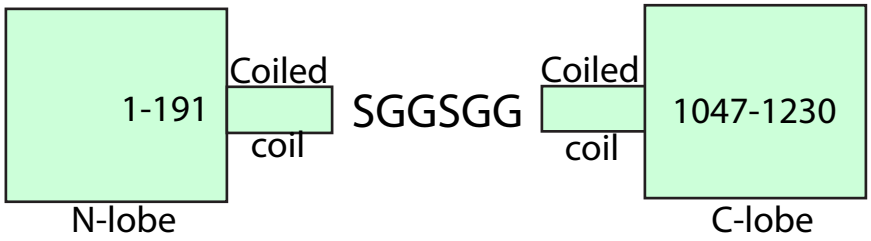


Figure 5-8: Smc3-ATPase construct used for biolayer interferometry experiments

The Smc3 N- and C- terminal globular domains each with short adjacent coiled-coil sequences were joined using a short linker consisting of the sequence SGGSGG.

The FP assay showed that the binding affinity of the WAPL domain with acetylated WIS1 peptide was halved compared to that with the wild type peptide. To verify this result, an acetyl-mimicking mutant of Smc3-ATPase carrying the mutations K112N, K113N (ScSmc3-ATPase$^{K112N, K113N}$) was made, as the protein appeared to be refractory to *in vitro* acetylation by Eco1 when tested. This observation was consistent with previous reports (for example, Borges et al. 2010). The acetyl-mimic mutant Smc3 has been used in previous studies and was shown to be an adequate substitute of the acetylated protein (Ben-Shahar et al., 2008). Compared to the wild-type proteins, an approximate 2-fold decrease in the affinity (K_d=88nM) between the proteins ScSmc3-ATPase$^{K112N, K113N}$ and ScWpl1$^{259-647}$ was observed.

Next to be tested was the effect of mutations in the WIS2-Wpl1 interface, based on the structure of the complex. In order to do this, binding between wild type ScWpl1$^{259-647}$ with ScSmc3-ATPase mutants - L163A/L166A/D168A and L163D/L166D, were tested. The dissociation constants obtained (K_d of 104 and 102 respectively) showed a weaker binding in both the cases. Alternatively, mutations L532E, L577R and N493E/N622E in the WAPL domain, with dissociation constants 82, 77 and 126 nM respectively, also resulted in a weakening of its interaction with the wild type Smc3-ATPase. Interestingly, a conserved lysine, K486 (K412 in AgWpl1), projecting out from Motif III, when mutated to asparagine, results in decreased affinity between the proteins (K_d=85nM) although it does not directly contact WIS2.

Table 5-5: Binding affinities between *S. cerevisiae* WAPL and Smc3 ATPase domains

Proteins	**K_d** (nM)	**K_{on}** ($1/Ms \times 10^3$)	**K_{off}** ($1/s \times 10^{-4}$)	**R^2**
wt-ScSmc3 ATPase / wt-ScWpl1$^{259-647}$	47	5.22 ± 0.07	2.48 ± 0.08	0.996
ScSmc3 K112N/K113N	88	4.88 ± 0.06	4.31 ± 0.11	0.993
ScSmc3 L163A/L166A D168A	104	4.77 ± 0.06	4.97 ± 0.12	0.993
ScSmc3 L163D/L166D	102	5.88 ± 0.08	5.97 ± 0.12	0.989
ScWpl1$^{259-647}$ L532E	82	4.20 ± 0.03	3.44 ± 0.06	0.998
ScWpl1$^{259-647}$ L577R	77	5.54 ± 0.06	4.28 ± 0.10	0.993
ScWpl1$^{259-647}$ N493E/N622E	126	3.92 ± 0.05	4.92 ± 0.09	0.995
ScWpl1$^{259-647}$ K486N	85	4.75 ± 0.05	4.05 ± 0.01	0.993
ScWpl1$^{259-647}$ D271A/D272A	167	2.37 ± 0.03	3.95 ± 0.06	0.998
ScWpl1$^{259-647}$ K376Q/K377Q/K382Q/K383Q	200	2.80 ± 0.04	5.60 ± 0.07	0.997

[experiments were performed with mutations in any one of the component proteins (either Wpl1 or Smc3) and only the component carrying mutation(s) has been indicated for clarity]

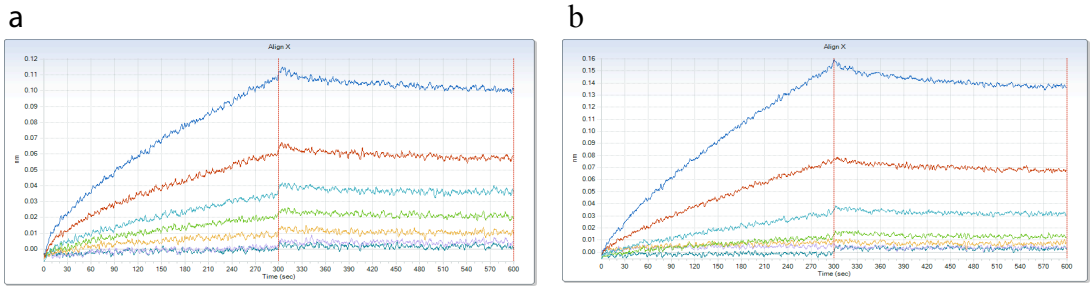


Figure 5-9: Representative binding data from biolayer interferometry

Typical association and dissociation curves obtained in case of (a) wild type ScWpl1$^{259-647}$ and ScSmc3-ATPase and (b) ScWpl1$^{259-647}$ D271/272A mutant and ScSmc3-ATPase. The curves in different colours represent the varying concentrations of ScWpl1$^{259-647}$ used (dark blue – 1200 nM, red – 800 nM, light blue – 400 nM, light green – 200 nM, orange – 100 nM, purple – 50 nM and dark green – 25 nM of the protein).

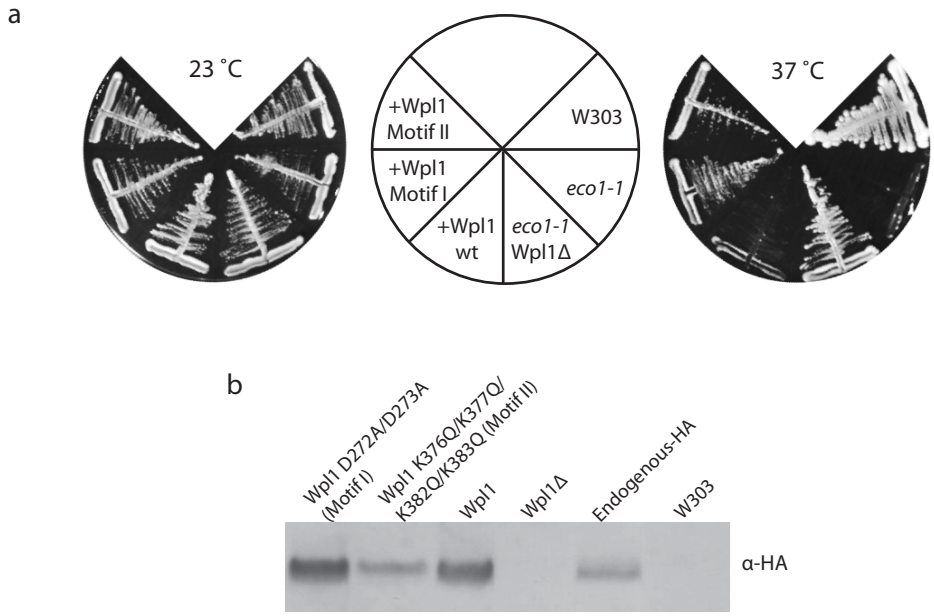


Figure 5-10: ***in vivo*** **validation of Wpl1-Smc3 interaction**

(a) Growth of different yeast strains at permissive (23 °C) and restrictive temperatures (37 °C). The *ts* mutant strain *eco1-1* fails to grow whereas deletion of the endogenous *WPL1* gene (indicated as eco1-1/Wpl1Δ strain) restores viability in these cells at the restrictive temperature. Introduction of a chimaeric wild-type *WPL1* gene through the plasmid yIPLac204 plasmid (shown as "+Wpl1wt") causes the cells to revert back to the lethal phenotype but introduction of a *WPL1* gene carrying mutations in either Motif I (shown as "+Wpl1 Motiff I") or in Motif II (shown as "+Wpl1 Motif II") does not cause lethality. The strain W303 was used as a control **(b)** Western blots showing the expressions of the Wpl1 protein in the respective strains detected using an α-HA antibody.

Finally, the effect of mutations in the conserved residues of Motifs I and II of the WAPL domain were verified. The ScWpl1259-647 mutants D271A/D272A (Motif I) and K376Q/ K377Q/ K382Q/ K383Q (Motif II) bound with significantly decreased affinities to wild-type ScSmc3-HD (Kd of 167 and 200 nM respectively). This probably shows that the WAPL domain interacts with Smc3 most strongly through its N-terminal conserved motifs in the cellular context.

5.4.2 *in vivo* functional assays

For these experiments, the temperature sensitivity of an *ECO1* mutant strain, called *eco1-1*, that can grow at 23ºC (permissive temperature) but not at 37ºC (restrictive temperature), was made use of. In the *eco1-1* strain background, the endogenous *WPL1* gene was replaced with a C-terminal 3X HA-tagged *WPL1*, which was transformed into

these cells after cloning into the plasmid pYIPLac204. Deletion of the *WPL1* gene from the *eco1-1* mutants allows the cells to grow normally at the restrictive temperature while replacing back the gene through transformation restores the temperature sensitive phenotype. However, if a defective *WPL1* gene (e.g., the mutants described previously) is introduced, then the strain retains the *WPL1* deletion phenotype, i.e., normal growth at 37ºC.

This assay was used to test the effect of all the Wpl1 mutants described before (as listed in Table 4-4 and 5-4). It was observed that Motif I (D271A/D272A) and Motif II (K376Q/K377Q/K382Q/K383Q) mutants were able to grow normally (Figure 5-8) and thus appeared to affect the functioning of Wpl1 the most. These results were obtained consistently in several trials but a similar impact was not observed for mutations in the other motifs, although some mutations, e.g., K486N, appeared to have a mild but inconsistent effect. Nevertheless, the results obtained from the *in vivo* experiments were in very good agreement with the biolayer interferometry based binding studies. Together, these results show that Smc3 probably binds most strongly to the conserved Motifs I and II of the WAPL domain, as mutations in these regions cause maximum disruption of the interaction between WAPL and Smc3 leading to severe cellular defects.

Chapter 6. Analyses of Wpl1-Pds5 interaction

Wpl1 is known to form a subcomplex with the HEAT repeat protein Pds5 and this interaction has been shown to be important for sister chromatid cohesion (Kueng et al., 2006). The interaction between the two proteins has been better characterized with the human orthologues of the respective proteins. The human Pds5 interacts with conserved FGF motifs present in the N-terminal of the corresponding WPl1 orthologue (Figure 1-3). However, the sequence conservation of the FGF motifs is restricted to the vertebrate Wpl1 orthologues and the lower eukaryotes including *Drosophila*, *C. elegans* and fungi, does not show the presence of similar motifs in their respective Wpl1 proteins. So, the interaction between the *A. gossypii* Pds5 and Wpl1 orthologues were analysed in detail in order to check whether the nature of the interaction is conserved among higher and lower eukaryotes. The experiments carried out for this analysis have been described in the present chapter.

6.1 Characterisation of the Wpl1-Pds5 complex

6.1.1 Reconstitution of the subcomplex

Interaction studies between AgPds5 and full-length AgWpl1 proteins were carried out to first reconstitute the subcomplex with purified proteins. For this experiment, the proteins were dialysed into a buffer containing 40 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT and were incubated together on a rocking platform for 2 hours at 4 °C. The proteins were then concentrated and put through an analytical size exclusion column and the elution peaks were analysed using a SDS-PAGE gel.

A stable complex could be observed between the AgWpl1 and AgPds5 wild-type proteins which elutes from an analytical size exclusion column (Superose 6; 24 mL bed volume) at about 13.8 mL as a sharp peak while any left-over Wpl1 elutes at approximately 15.5 mL.

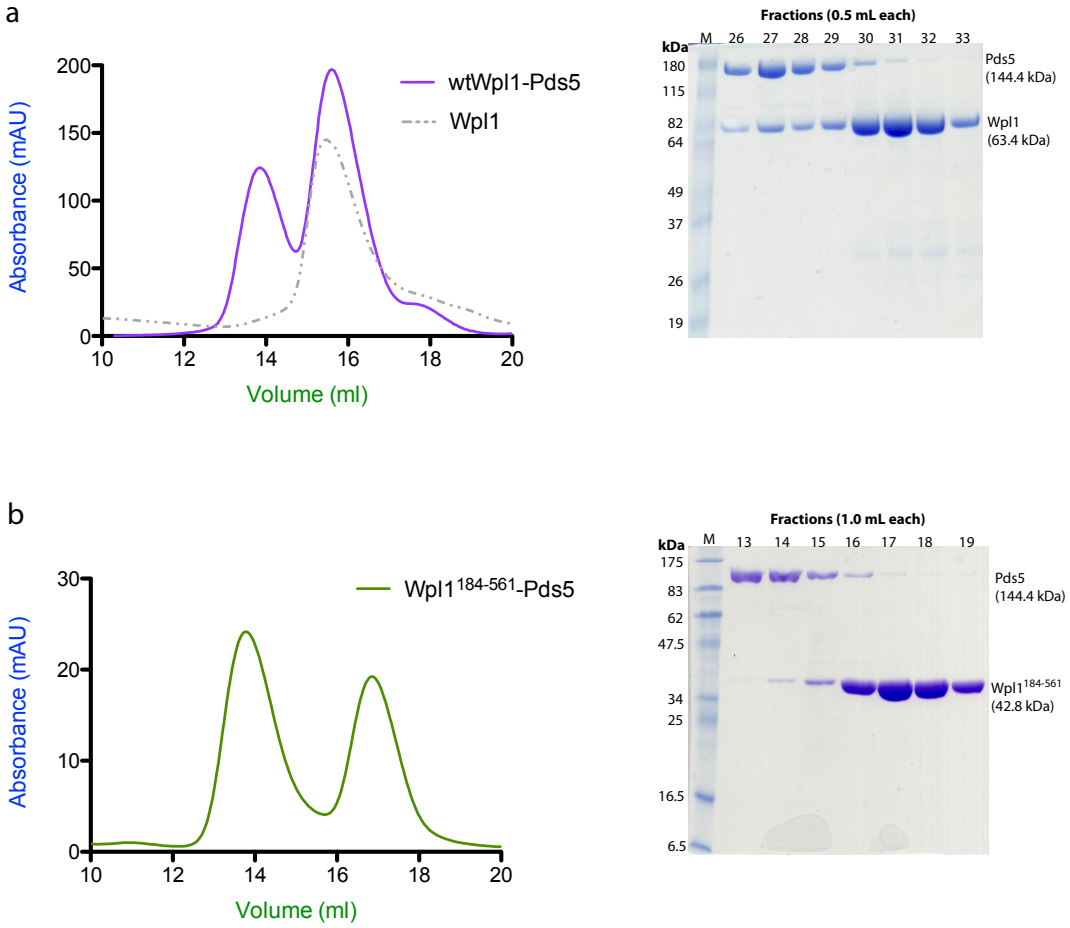


Figure 6-1: Reconstitution of Wpl1-Pds5 subcomplex

(a) Gel filtration trace and SDS-PAGE showing the stable complex formed between AgWpl1 (full-length) and AgPds5 which elute together around 14 mL while the free AgWpl1 elutes out at 16 mL from a Superose6 size-exclusion column. (b) Similar experiment performed with AgWpl1$^{184\text{-}561}$ shows that the WAPL domain is not capable of forming a similar complex with Pds5. The respective proteins elute out of a size-exclusion column as separate peaks.

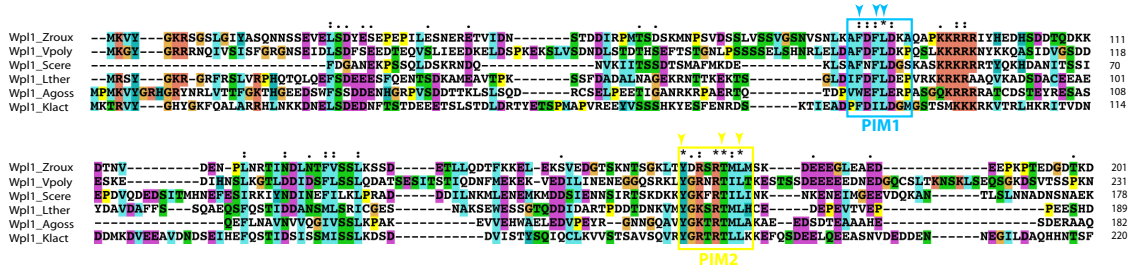


Figure 6-2: **Alignment of Wpl1 N-terminal sequences**

Multiple sequence alignment using only the N-terminal sequences of Wpl1 orthologues from *Z. rouxii*, *V. polyspora*, *S. cerevisisae*, *L. thermotolerans*, *A. gossypii* and *K. lactis*. The positions of the PIM1 (blue box) and PIM2 (yellow box) motifs are indicated. The residues mutated in each motif have been pointed out.

A similar experiment was also performed with purified AgWPl1$^{184-561}$ (WAPL domain only) and AgPds5 proteins. However, in this case, the complex could not be observed and the size-exclusion peaks corresponding only to the individual proteins were obtained. This experiment confirmed that the N-terminal interactions of the Wpl1 protein are conserved and as in human, the WAPL domain of fungal Wpl1 orthologues do not participate in its interaction with Pds5.

6.1.2 N-terminal sequence analysis of Wpl1

Analysis of the fungal Wpl1 orthologues based on their sequence alignment shows a few conserved regions in the N-terminal of the proteins. Two of these identified stretches are especially conserved which is mainly contributed by the hydrophobic residues present in the respective motifs and thus appeared to be similar to the FGF motifs present in the human Wapl. In the *A. gossypii* Wpl1 sequence, these motifs span residues 79-86 and 152-161 (PIM1 and PIM2; ***P***ds5 ***i***nteracting ***m***otif). The PIM1 motif contains two conserved phenyalanines (first one replaced with tryptophan W80 in AgWpl1) and a totally conserved leucine whereas PIM2 consists of the totally conserved residues tyrosine, arginine and threonine followed by a conserved hydrophobic four-residue stretch (Figure 5-2).

6.1.3 Design and interaction of N-terminal Wpl1 mutants

The PIM1 and PIM2 motifs, owing to their hydrophobicity and sequence conservation, could be involved in the interaction of Wpl1 with Pds5. Mutations in these motifs of AgWpl1 were designed in order to test their effect on the Wpl1-Pds5 interaction. Two Wpl1 mutants – (a) PIM1 mutant with W80, F82 and L83 residues replaced by alanines and (b) PIM2 mutant in which residues Y153, T158 and L160 were mutated to alanines were made. This allowed verification of the role of these conserved stretches in the Wpl1-Pds5 interaction. When tested, it was observed that mutations in PIM1 result in a significantly weakened interaction between Wpl1 with Pds5 as the gel filtration profile showed a single major peak corresponding to Wpl1 alone (16 mL) but

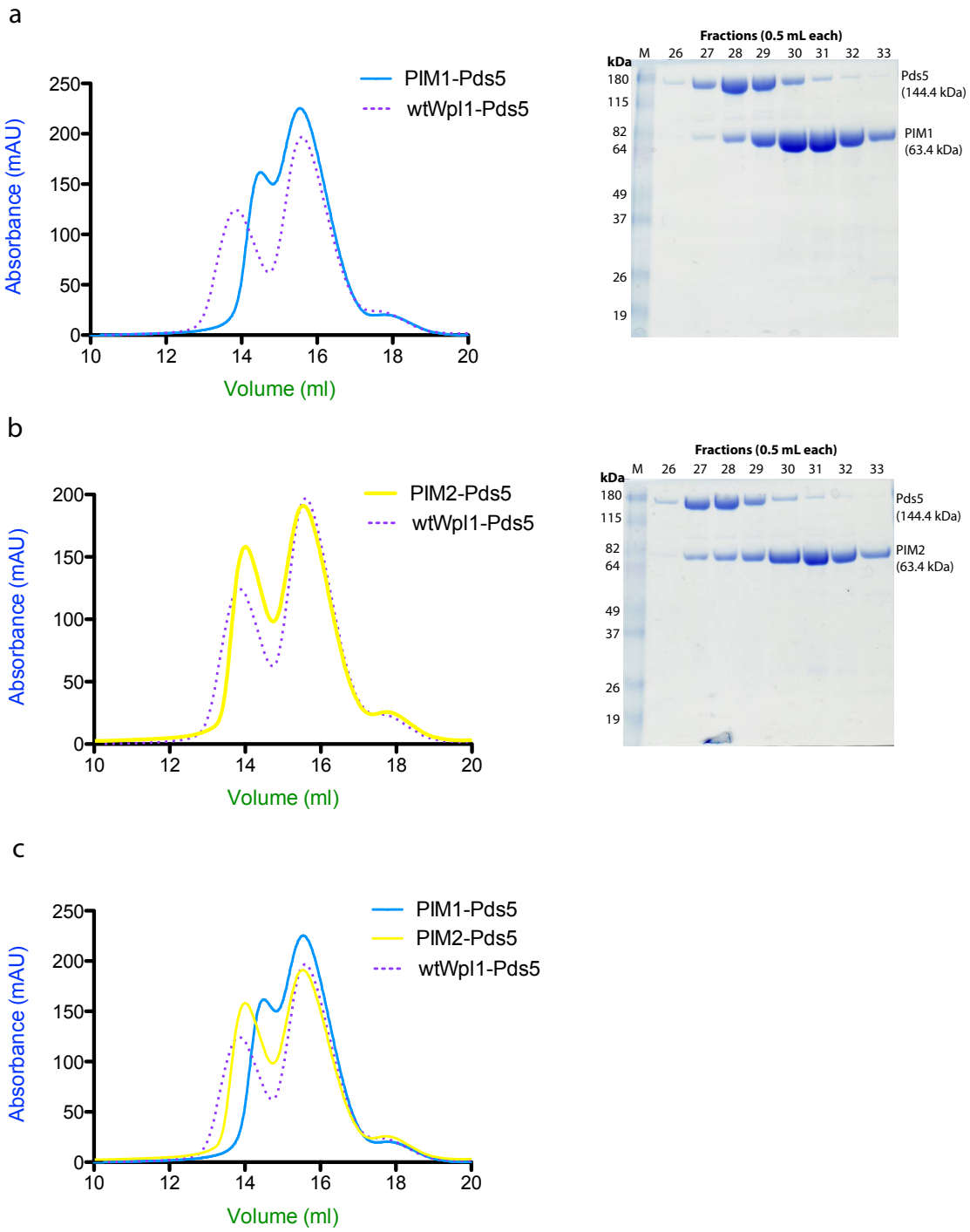


Figure 6-3: Interaction of N-terminal Wpl1 mutants with Pds5

Gel filtration traces and SDS-PAGE gel profiles showing (a) disruption of interaction between Pds5 and Wpl1mutations in PIM1 motif (denoted as PIM1), (b) no significant disruption of the interaction between the PIM2 mutant and Pds5 and (c) traces of both PIM1 and PIM2 superimposed for comparison. PIM1-Pds5 trace is shown in blue, PIM2-Pds5 in yellow. The wild-type Wpl1 Pds5 trace is shown as a purple dotted plot for reference.

with a small shoulder. The gel filtration peak profile of interaction between the PIM2 mutant Wpl1 and Pds5 was not much different compared to that of wild-type Wpl1 and Pds5 although it showed a very small (~0.25mL) shift of the complex peak (Figure 6-3). When analysed on a SDS-PAGE gel, the effect of PIM1 mutations were quite evident as Pds5 and the PIM1 mutant proteins appeared separated out. However, PIM2 mutations did not appear to disrupt the interaction much although there was slight difference (a shift of the complex peak toward the right) compared to the wild-type interaction profile.

The results show that the PIM1 motif of fungal Wpl1 orthologues might be the key motif mediating its interaction with Pds5. Furthermore, the residues W80 (replaced by F in all other fungal orthologues) and F82 might recapitulate the effect of the FGF motifs of the human Wpl1. Finally, the effect of the PIM2 mutant on the interaction, though small, shows that the interaction interface might be more extended and might result from multiple contacts between the respective proteins which in Wpl1 is restricted to the N-terminal sequence.

Chapter 7. Discussion

Sister chromatid cohesion is a basic determinant of accurate chromosome segregation while the cohesin complex additionally influences processes like gene transcription and expression, thereby playing a critical role in organismal development. These varied roles make cohesin one of the most intriguing macromolecular complexes. As can be expected for any key physiological assembly, activities of cohesin are tightly regulated, starting right from the loading of the complex onto DNA through establishment and maintenance of cohesion to the eventual removal from chromatin upon fulfilment of its role. Regulation at each level is integral to the functioning of the complex while engaging a distinct set of highly conserved factors. Loading of cohesin, occurring during telophase in vertebrates and late G1 in yeast depends on Scc2/Scc4. Activation of the complex during S phase is distinct from its loading onto chromosomes and is catalysed by the acetyltransferase Eco1 following which sister chromatids are entrapped inside the cohesin ring. In vertebrates, the protein sororin plays an important role in maintenance of cohesion during G2 phase. The prophase pathway, which initiates removal of cohesin from vertebrate chromatin is brought about mainly by the protein Wapl while the complete dissolution of cohesion, triggered by the cleavage of centromeric cohesin is mediated by the protease separase during the onset of anaphase. Budding yeast differs from vertebrates in lacking a prophase pathway and separase-mediated cleavage accounts for the removal of both centromeric as well as arm cohesin. The Wapl orthologue in yeast, known as Wpl1 or Rad61 has a proposed role in influencing cohesion establishment by maintaining cohesin in a state that is unable to stably interact with or entrap DNA prior to acetylation by Eco1.

Apart from a host of regulatory proteins, the nature of the subunits and the architecture of the cohesin core itself allow the complex to be intricately regulated. The dependence of the stability of Smc1/Smc3 heterodimer on ATP binding and hydrolysis and the flexible nature of the Smc monomers owing to the presence of extensive coiled coils are central to the dynamic nature of the association of cohesin with DNA. The subunit Scc1, besides being the target for proteolytic cleavage by separase, together with Scc3, is also the hub of regulation by a battery of kinases and phosphatases.

Cohesin and its regulators have been subjected to intense research and considerable progress has been made toward the identification and characterisation of the events leading to the activation as well as removal of cohesin. However, as is clear from the description of the process in earlier sections, a complete understanding of cohesion establishment is not possible without a clear demonstration of the mechanism of Wpl1 functioning. Although Wpl1 and its orthologues show a variable N-terminal region, they invariably consist of a C-terminal helical domain that is conserved across eukaryotes. Given the conservation, the apparently variant mechanisms by which the lower and higher eukaryotic orthologues of this protein function was baffling. Therefore, an attempt has been made in the work described in this thesis, to reveal the conserved aspects of Wpl1/Wapl functioning through elucidation of the crystal structure and identification and characterisation of the interactions mediated by the conserved C-terminal domain of the protein.

7.1 The WAPL domain structure reveals novel features

The WAPL domain of the budding yeast Wpl1 protein failed to yield high-quality crystals, following which the WAPL domain of the *Ashbya gossypii* orthologue was crystallised allowing high-resolution structure solution. The WAPL domain, as predicted earlier (Kueng et al., 2006), showed an α-helical repeat structure with an elongated morphology. Nonetheless, the structure revealed a number of novel and intriguing features, which provided a framework for the biochemical characterisation of the protein that was carried out in this work. The overall fold of the WAPL domain resembles that of the canonical HEAT repeats, domains known to mediate protein-protein interactions. The WAPL structure, however, can be divided further into three subdomains, comprising of the N- and C-terminal lobes and the middle HEAT region, a feature that makes this structure unique. In addition, the protrusion of a conserved helix and an adjacent loop rich in basic residues sets the WAPL domain further apart from related structures.

The structure also helped identify conserved surface motifs, two of which formed an acidic (Motif 1) and a basic patch (Motif 2) on the N-terminal of the domain. Mutation in a conserved aspartate, D272 (to Gly), in budding yeast Wpl1 was previously shown to suppress lethality in *ECO1* mutant cells (Rowland et al., 2009). The residue corresponds to D206 in *A. gossypii* Wpl1, located within the conserved acidic patch of the domain. The phenotype associated with the mutation goes on to show the importance of this conserved patch and this was also confirmed by the *in vivo* results presented in this work (the suppression of lethality at restrictive temperature of *eco1-1* cells carrying the mutations D271A/D272A in Wpl1). Charge neutralising mutations in the basic patch (Motif 2), consisting of the protruding helix and an extended adjoining loop, also resulted in strong phenotypes in the *in vivo* assays showing the importance of this conserved region as well in Wpl1 functioning. These results, taken together, show the N-lobe, with its conserved and charged surface, is absolutely essential for the functioning of Wpl1.

The structure also revealed additional conserved stretches, named Motifs 3, 4 and 5, which together form a hydrophobic patch on the C-terminal surface of the WAPL domain. Parts of these three motifs together form a conserved pocket with a core formed exclusively of hydrophobic residues. However, mutating residues in this hydrophobic region of the domain did not cause as drastic effects as were observed in case of Motifs 1 and 2 when tested by the *in vivo* assay.

7.2 Assigning a function to the WAPL domain

The previously identified binding partners of the human orthologue of Wpl1 included Pds5 and Scc3, although both of these were shown to interact with the variable N-terminal (Rowland et al., 2009) and (Shintomi and Hirano, 2009). An analysis of the interaction between the Wpl1 and Pds5 proteins from *A. gossypii*, described in this work, also showed the interaction to be mediated through the N-terminal of Wpl1 similar to that seen in humans. These observations suggested a role of the WAPL domain separate from that of the N-terminal of the protein. In the search for binding partner(s) of the WAPL domain, the peptide array screens identified the Smc3 ATPase head and Scc1

proteins as the probable interactors. The peptide hits of Smc3, named WIS1, WIS2 and WIS3, were further validated using fluorescence polarisation assays. In addition, the interaction between the WAPL domain and an intact Smc3-ATPase head was also confirmed using the biolayer interferometry assay, which demonstrated a strong binding between the two (K_D of ~47 nM).

All the three peptide hits could be mapped onto surface exposed regions on an Smc3-ATPase model, which was based on the homologous Smc1 head structure and surprisingly, the WIS1 peptide harboured the two conserved Eco1 target lysines. The location of the peptides WIS2 and WIS3 in the Smc3-ATPase model is also worth noting – WIS2 forms a pair of helices adjacent and directly above the conserved Q-loop (refer to section 1.3.5) while WIS3 constitutes the base of the coiled coil emanating from the globular head. The Q loop, as described before, links ATP hydrolysis to conformational changes in the trans-membrane domain in ATPase transporters. Significantly, a comparison of the Smc ATPase with the recently reported structures of Rad50 ATPase (containing short coiled coil stretches) in complex with the Rad50-binding domain of Mre11 (Mre11-RBD) ((Williams et al., 2011); Fig. 7-1) showed that the WIS2 sequence corresponds to the so-called signature coupling helices (SCH) of Rad50 ATPase. The pair of signature coupling helices in Rad50, also placed adjacent to the corresponding Q loop, undergoes a conformational change upon the binding of nucleotide to the protein. This change in conformation in the SCHs causes a shift in the position of the coiled-coils with respect to the nucleotide-binding (ATPase) domain and as a result of this movement, the Mre11-RBD positioned at the base of the Rad50 coiled-coil is also relocated thus bringing the protein in proximity to its DNA substrate. Thus, the SCHs have a clear role in mediating an intramolecular cross talk with functional implications and given the fact that these ATPases act via highly conserved mechanisms, the observation that WIS2 represents the SCHs, is quite significant. WIS3, as mentioned before, corresponds to the base of the coiled-coil arising from the C-terminal lobe of the Smc ATPase and again seems to be functionally relevant as this region is prone to conformational changes based on the Rad50-Mre11 structure. The WAPL domain therefore appears to be perfectly positioned so as to be able to regulate

functioning of Smc3 by influencing the ATPase activity or/and the conformation of the protein.

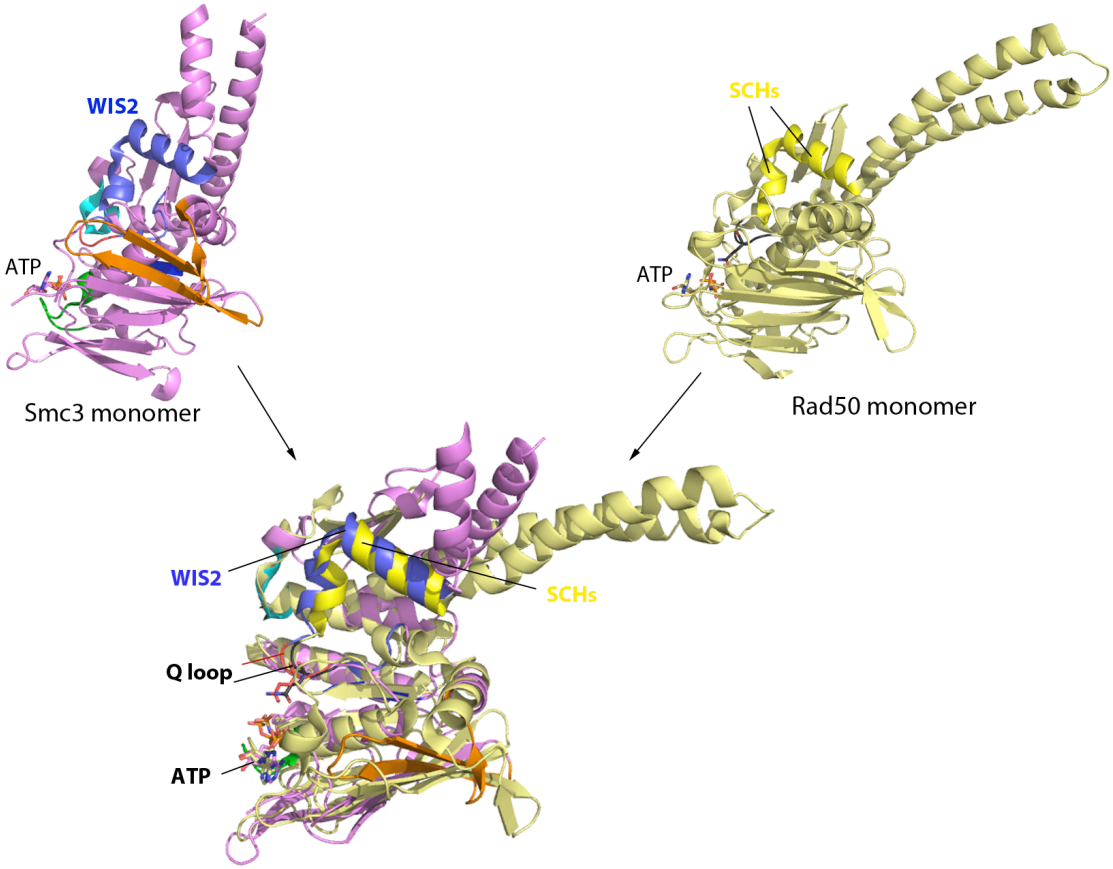


Figure 7-1: Structural alignment of Smc and Rad50 ATPases

The structures of monomeric Smc and Rad50 ATPases (top) are aligned (below). The WIS2 helices (shown in violet) correspond to the SCHs (bright yellow) in the aligned structures. The positions of the Q loop and the ATP have also been marked.

Coincidentally, a co-crystal structure of the WAPL domain in complex with the WIS2 peptide could be obtained. This not only re-confirmed the interaction between WAPL domain and the WIS2 region of the Smc3 but also helped to map the putative biding site of WIS2. Interestingly, the peptide binds to the conserved hydrophobic surface on WAPL, with two of the conserved leucines of WIS2 stacking against residues at the base of the pocket located within the patch. Together with the mutagenesis data, this complex structure enabled to create a model of WAPL binding on the Smc3 head (see below).

The interaction between the WAPL domain and the Smc3-ATPase was verified in context of the proteins by introducing structure-guided (based on both WAPL and WAPL-WIS2 structures) mutations in conserved residues of both the partners. An exhaustive number of mutations were introduced in the WIS2 region of Smc3-ATPase and its corresponding binding surface on WAPL. Although these mutations in the WIS2 interface (refer to Section 5.4.1 and Table 5-5) led to a compromised binding between the proteins, the decrease was only about two-fold in most of the cases. Also, the mutations when tested in vivo did not show a marked phenotype (do not suppress lethality in *eco1-1* mutant cells).

Based on the WAPL-WIS2 complex structure, the position of the WAPL domain on Smc3-ATPase model can be tentatively traced as shown in Fig. 7-2. According to this model, the distal C-terminal of the WAPL domain is positioned above the WIS1 sequence of Smc3 and the N-terminal (containing Motifs I and II constituting the conserved charged patches) of the domain is appended near to the base of the coiled-coil, that is, the WIS3 sequence of Smc3. Results of mutagenesis experiments show that charge neutralising mutations in Motifs I and II cause a four to five-fold decrease in affinity *in vitro* and drastic phenotypes *in vivo* corresponding to considerable suppression of lethality in *eco1-1* mutant cells. These observations/results together indicate that the N-terminal of the WAPL domain interacts with the WIS3 region of Smc3-ATPase and that this interaction is much stronger compared to the interaction at the WIS2 interface. Consequently, mutations in the WIS3 interface (in residues of Motif I and II) results in a complete or near complete disruption of the Wpl1 binding to Smc3. The mutations therefore result in a non-functional Wpl1, which lead to the strong phenotypes that are observed in the *in vivo* experiments. The length of the WAPL domain (~85 Å) is much larger compared to the width of Smc3 (~35 Å) and the distance between WIS1 and WIS3 sequences (~50 Å). However, in the actual binding mode, the WAPL domain and possibly the Smc3 might undergo major conformational changes so as to engage all the binding sites, WIS1, WIS2 and WIS3 together (Fig. 7-2). Also, from these results, the WAPL domain seems to be tethered strongly through its N- and C-termini to the Smc3 and this might be the reason why mutations in the WIS2 interface

(which is approximately in the middle of the WAPL domain) do not cause drastic effects either *in vitro* or *in vivo*.

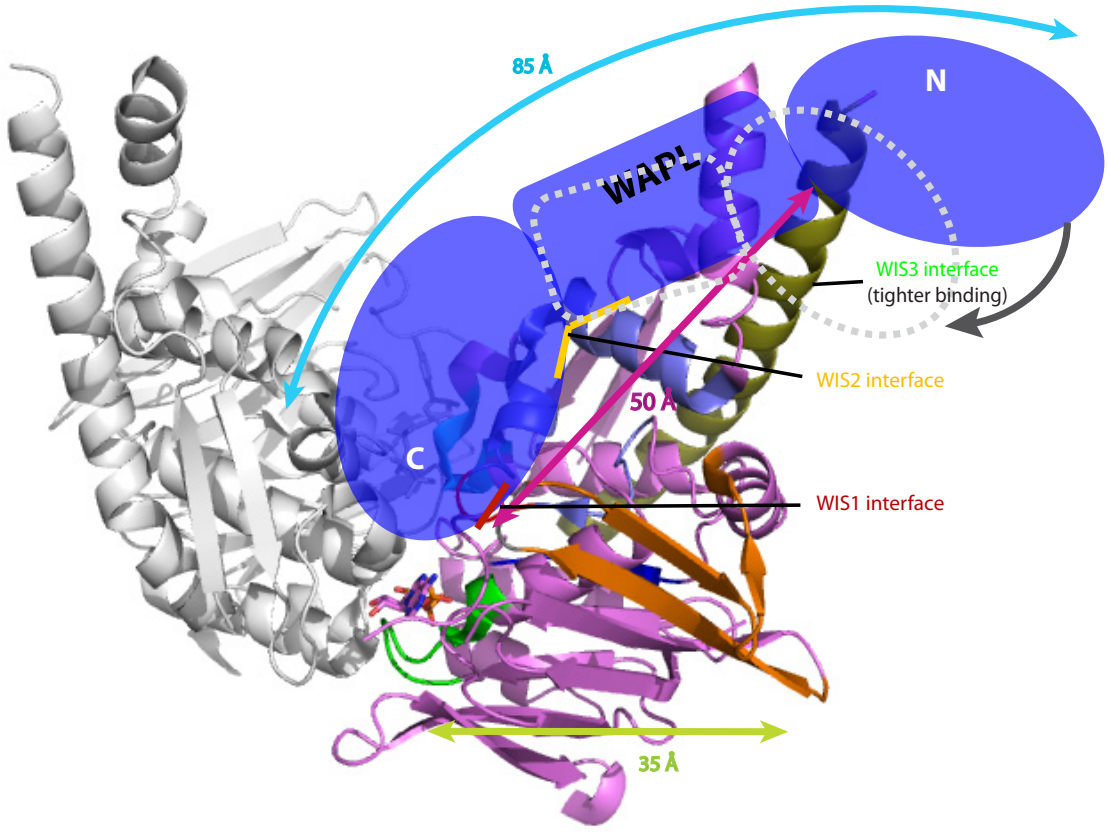


Figure 7-2: Model of WAPL binding to Smc3 ATPase

In this model the C-terminal of the WAPL domain (shown as a cartoon in blue) is probably positioned near the WIS1 sequence of Smc3 while the N-terminal is directed towards the coiled-coil of Smc3 containing the WIS3 sequence at its base. The actual binding might involve major conformational changes either only in the WAPL domain or in both WAPL and the Smc3. The N-lobe of the WAPL domain probably wraps around (depicted as grey dotted outline) the Smc3 coiled-coil in order to bind to the WIS3 sequence. The lengths of the WAPL domain (in light blue) and the Smc3 ATPase head (diagonal in pink and width in green) have been shown to provide an idea of the relative dimensions of the proteins.

In addition to the interaction with Smc3 ATPase, preliminary results based on peptide array (Section 5.1) showed that the WAPL domain also binds the cohesin subunit Scc1. This interaction obviously needs further verification using intact proteins, but if WAPL really does interact with Scc1, it might lead to key insights into the mechanism of Wpl1 functioning. Scc1 forms two crucial junctions, the Smc1-Scc1 and the Scc1-Smc3 gates, one or both of which might act allow DNA exit. Therefore, interaction of Wpl1 with Scc1 might directly enable control of these exit gates.

7.3 An explanation for opposing actions of Wpl1 and Eco1

Since the WAPL domain was found to interact with the Smc3-ATPase and one of the target sites was found to contain the conserved pair of lysines acetylated by Eco1 (the WIS1 sequence), an obvious question was whether acetylation of Smc3 affected the interaction. To address this, the binding of WAPL was checked first with an acetylated WIS1 peptide and subsequently with an acetyl-mimic Smc3-ATPase mutant (K112N/K113N). Compared to the wild-type counterparts, the acetylated peptide as well as the mutant protein showed a two-fold decrease in the affinity to the WAPL domain. A reason for this marginal decrease in affinity may be that the asparagine side chain cannot fully mimic an acetyl lysine and therefore does not weaken the interaction as much as an acetyl group sitting on a lysine would. Nonetheless, a considerable weakening in the binding of WAPL to the Smc3 might result in a dislocation of WAPL from Smc3, leading to a complete dissociation of the protein from the cohesin complex. This outcome is definitely not desirable and does not agree with previous studies showing Wpl1 to be constitutively associated with cohesin (Kueng et al., 2006). Moreover, if Wpl1 regulates the Smc3 ATPase activity through the SCHs, a slight change in the conformation of the helices, as a consequence of a partial weakening of the interaction between the proteins, might be enough to initiate functionally relevant changes in the cohesin complex (Fig. 7-3). Therefore, it is tempting to think that the two-fold decrease in affinity observed is physiologically relevant.

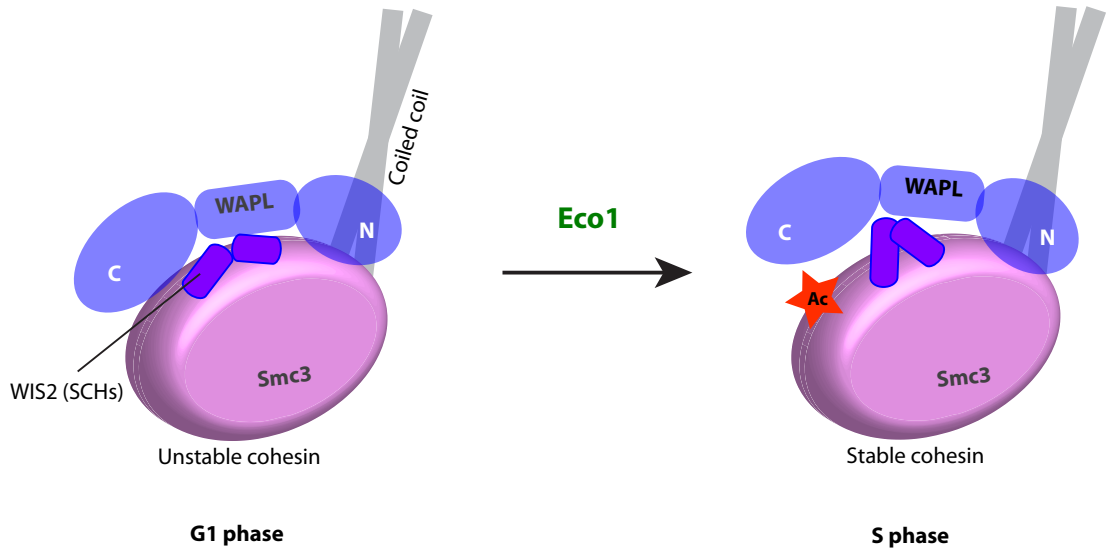


Figure 7-3: Acetylation-mediated stabilisation of cohesin

After cohesin is loaded during G1 phase, it is associates with DNA in an unstable manner owing to a stronger binding of WAPL with the Smc3 ATPase (left). During DNA replication in S phase, acetylation of Smc3 head weakens WAPL binding and causing a displacement of the protein. This displacement causes a conformational change in the WIS2 helices (or SCHs) leading to the stabilisation of cohesin.

7.4 Conclusions

The human orthologue of Wpl1 was shown to maintain cohesin in a dynamic state such that the resident time of the complex on DNA is low. Due to this reason, there is continuous turnover in the chromatin bound population of the cohesin until acetylation of Smc3 by Eco1 during S phase. In yeast, it was found that deletion in the *WPL1* gene allows the cells to overcome the need for Eco1 and therefore acetylation based stabilisation of cohesin-DNA interaction is no longer required. In addition, a recent study has shown that the yeast Wpl1, similar to the human orthologue, increases turnover of cohesin on yeast chromatin. Interestingly, the study also showed that Wpl1 promotes release of DNA through the Smc3-Scc1 junction but the exit gets sealed once Smc3 is acetylated (Chan et al., 2012). These results clearly show a conserved function of Wpl1 in eukaryotes.

In spite of all these observations and insights, it was still not understood how Wpl1 exerts its effect on cohesin. So, it was not clear whether the protein plays an indirect

role in destabilising cohesin-DNA interactions (maybe through known proteins like Pds5 or some other unknown factor) or if it has a more direct effect. The present work provides insights into the mechanism of functioning of Wpl1 by demonstrating a direct interaction between Wpl1 and cohesin and has also provided explanations as to how acetylation of Smc3 might help in the establishment of cohesion. Furthermore, the results imply that Wpl1 is positioned on the Smc3 head in a way that it might be able to regulate its ATPase activity and therefore the associated conformational changes in cohesin. Although the results presented here appear promising and address some of the longstanding issues, it is equally important to note that cohesion is an intricate process and an exhaustive understanding of regulation by Wpl1 and Eco1 obviously needs further work. It might as well be possible that the mechanisms suggested in this study act in conjunction with other processes to regulate cohesin.

Another related but speculative aspect of sister chromatid cohesion is the reason for maintaining unstable cohesin-DNA interaction prior to the S phase. One reason for the unstable association may be that if cohesin is allowed to stably entrap DNA as soon as it loads onto the chromosomes, then it would be difficult for the replication machinery to pass through this 'rigid' cohesin ring during DNA replication. On the other hand, if cohesin is maintained in an unstable or dynamic state, the passage of the replication machinery can take advantage of this unstable state of the complex. This might be the reason why cohesin-DNA interaction is stabilised in a replication fork dependent manner by the fork associated Eco1. In addition, stable cohesin-DNA interactions earlier than S phase might result in inter-chromosome linkages, a situation that can be avoided if cohesin becomes active only during replication. However, a definitive answer is still not available and awaits detailed and convincing tests.

7.5 Further work

Based on the results presented, an important finding has been that the WAPL domain binds the signature-coupling helices of Smc3 ATPase as well as at the base of the coiled-coil. This suggests a shift in the ATPase activity of Smc3 or the Smc1/Smc3 complex depending on Wpl1 bound or unbound (acetylated) states. This prediction

needs to be confirmed by checking the ATPase activities of ideally the Smc1/Smc3 full-length complex or at least with the heterodimeric ATPase heads of Smc1/Smc3, both in the presence and absence of Wpl1 (or WAPL domain). However, obtaining usable amounts of relatively pure dimeric Smc1/Smc3 either as full-length or as ATPase truncations might prove challenging. Trials to overexpress and purify the full-length heterodimer have so far been hampered due to no or extremely low yields and attempts to purify the heterodimeric head truncations have suffered due to instability of such a dimer.

The peptide array results and preliminary fluorescence polarisation assay data (not presented here) also suggest an interaction between the WAPL domain with the N-terminal of Scc1. It would be quite interesting to characterise this interaction further as well and especially so because of the finding that the Smc3-Scc1 junction acts as a DNA exit gate (Chan et al., 2012).

The reason for Wpl1 forming a stable complex with Pds5 is not yet clear. So, as a long-term objective, the effect of this complex on the ATPase activity of the Smc1/Smc3 heterodimer can also be verified and compared to the effect of Wpl1 alone on the complex. Also, it would be interesting to see if the complex interacts differently with Smc3 than Wpl1 alone does.

Finally, a structure of the Smc3 ATPase head alone, in complex with Smc1 ATPase or/and with WAPL domain would be ideal toward developing a deeper understanding of the underlying mechanisms of regulation of cohesion.

Reference List

Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M.A., and Nasmyth, K. (2001). Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. Cell *105*, 459-472.

Almedawar, S., Colomina, N., Bermudez-Lopez, M., Pocino-Merino, I., and Torres-Rosell, J. (2012). A SUMO-Dependent Step during Establishment of Sister Chromatid Cohesion. Current biology : CB.

Anderson, D.E., Losada, A., Erickson, H.P., and Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. The Journal of cell biology *156*, 419-424.

Arumugam, P., Gruber, S., Tanaka, K., Haering, C.H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Current biology : CB *13*, 1941-1953.

Bernard, P., Drogat, J., Maure, J.F., Dheur, S., Vaur, S., Genier, S., and Javerzat, J.P. (2006). A screen for cohesion mutants uncovers Ssl3, the fission yeast counterpart of the cohesin loading factor Scc4. Current biology : CB *16*, 875-881.

Birkenbihl, R.P., and Subramani, S. (1992). Cloning and characterization of rad21 an essential gene of Schizosaccharomyces pombe involved in DNA double-strand-break repair. Nucleic acids research *20*, 6605-6611.

Blat, Y., and Kleckner, N. (1999). Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. Cell *98*, 249-259.

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Rolef Ben-Shahar, T., and Uhlmann, F. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. Molecular cell *39*, 677-688.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta crystallographica Section D, Biological crystallography *59*, 2023-2030.

Chan, K.L., Roig, M.B., Hu, B., Beckouet, F., Metson, J., and Nasmyth, K. (2012). Cohesin's DNA Exit Gate Is Distinct from Its Entrance Gate and Is Regulated by Acetylation. Cell *150*, 961-974.

Chen, J., Lu, G., Lin, J., Davidson, A.L., and Quiocho, F.A. (2003). A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Molecular cell *12*, 651-661.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Molecular cell *5*, 243-254.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica Section D, Biological crystallography *62*, 1002-1011.

Davis (1971).

.

DiNardo, S., Voelkel, K., and Sternglanz, R. (1984). DNA topoisomerase II mutant of Saccharomyces cerevisiae: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. Proceedings of the National Academy of Sciences of the United States of America *81*, 2616-2620.

Dorsett, D., Eissenberg, J.C., Misulovin, Z., Martens, A., Redding, B., and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development *132*, 4743-4753.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica Section D, Biological crystallography *66*, 486-501.

Evans, P. (2006). Scaling and assessment of data quality. Acta crystallographica Section D, Biological crystallography *62*, 72-82.

Farcas, A.M., Uluocak, P., Helmhart, W., and Nasmyth, K. (2011). Cohesin's concatenation of sister DNAs maintains their intertwining. Molecular cell *44*, 97-107.

Furuya, K., Takahashi, K., and Yanagida, M. (1998). Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G1 phase. Genes & development *12*, 3408-3418.

Game, J.C., Birrell, G.W., Brown, J.A., Shibata, T., Baccari, C., Chu, A.M., Williamson, M.S., and Brown, J.M. (2003). Use of a genome-wide approach to identify new genes that control resistance of Saccharomyces cerevisiae to ionizing radiation. Radiation research *160*, 14-24.

Gillespie, P.J., and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Current biology : CB *14*, 1598-1603.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell *127*, 523-537.

Gruber, S., Haering, C.H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765-777.

Guacci, V., Hogan, E., and Koshland, D. (1994). Chromosome condensation and sister chromatid pairing in budding yeast. The Journal of cell biology *125*, 517-530.

Guacci, V., Hogan, E., and Koshland, D. (1997a). Centromere position in budding yeast: evidence for anaphase A. Molecular biology of the cell *8*, 957-972.

Guacci, V., Koshland, D., and Strunnikov, A. (1997b). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. Cell *91*, 47-57.

Haering, C.H., Lowe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Molecular cell *9*, 773-788.

Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., and Lowe, J. (2004). Structure and stability of cohesin's Smc1-kleisin interaction. Molecular cell *15*, 951-964.

Hartman, T., Stead, K., Koshland, D., and Guacci, V. (2000). Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. The Journal of cell biology *151*, 613-626.

Hauf, S., Waizenegger, I.C., and Peters, J.M. (2001). Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science *293*, 1320-1323.

Hirano, M., and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. The EMBO journal *21*, 5733-5744.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic acids research *38*, W545-549.

Hopfner, K.P., Karcher, A., Shin, D.S., Craig, L., Arthur, L.M., Carney, J.P., and Tainer, J.A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell *101*, 789-800.

Hoque, M.T., and Ishikawa, F. (2001). Human chromatid cohesin component hRad21 is phosphorylated in M phase and associated with metaphase centromeres. The Journal of biological chemistry *276*, 5059-5067.

Hornig, N.C., Knowles, P.P., McDonald, N.Q., and Uhlmann, F. (2002). The dual mechanism of separase regulation by securin. Current biology : CB *12*, 973-982.

Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K.L., Upcher, W., Godlee, C., Roig, M.B., Shirahige, K., and Nasmyth, K. (2011). ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. Current biology : CB *21*, 12-24.

Hung, L.W., Wang, I.X., Nikaido, K., Liu, P.Q., Ames, G.F., and Kim, S.H. (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. Nature *396*, 703-707.

Ishiguro, T., Tanaka, K., Sakuno, T., and Watanabe, Y. (2010). Shugoshin-PP2A counteracts casein-kinase-1-dependent cleavage of Rec8 by separase. Nature cell biology *12*, 500-506.

Ivanov, D., Schleiffer, A., Eisenhaber, F., Mechtler, K., Haering, C.H., and Nasmyth, K. (2002). Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion. Current biology : CB *12*, 323-328.
Kabsch, W. (2010). Xds. Acta crystallographica Section D, Biological crystallography *66*, 125-132.

Kenna, M.A., and Skibbens, R.V. (2003). Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. Molecular and cellular biology *23*, 2999-3007.

Kerrebrock, A.W., Miyazaki, W.Y., Birnby, D., and Orr-Weaver, T.L. (1992). The Drosophila mei-S332 gene promotes sister-chromatid cohesion in meiosis following kinetochore differentiation. Genetics *130*, 827-841.

Kim, H.S., Baek, K.H., Ha, G.H., Lee, J.C., Kim, Y.N., Lee, J., Park, H.Y., Lee, N.R., Lee, H., Cho, Y.*, et al.* (2010). The hsSsu72 phosphatase is a cohesin-binding protein that regulates the resolution of sister chromatid arm cohesion. The EMBO journal *29*, 3544-3557.

Kitajima, T.S., Hauf, S., Ohsugi, M., Yamamoto, T., and Watanabe, Y. (2005). Human Bub1 defines the persistent cohesion site along the mitotic chromosome by affecting Shugoshin localization. Current biology : CB *15*, 353-359.

Klein, F., Mahr, P., Galova, M., Buonomo, S.B., Michaelis, C., Nairz, K., and Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell *98*, 91-103.

Koshland, D., and Hartwell, L.H. (1987). The structure of sister minichromosome DNA before anaphase in Saccharomyces cerevisiae. Science *238*, 1713-1716.

Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A.*, et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nature genetics *36*, 631-635.

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., and Peters, J.M. (2006). Wapl controls the dynamic association of cohesin with chromatin. Cell *127*, 955-967.

Kurze, A., Michie, K.A., Dixon, S.E., Mishra, A., Itoh, T., Khalid, S., Strmecki, L., Shirahige, K., Haering, C.H., Lowe, J.*, et al.* (2011). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. The EMBO journal *30*, 364-378.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573-578.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. Acta crystallographica Section D, Biological crystallography *62*, 48-57.

Lipp, J.J., Hirota, T., Poser, I., and Peters, J.M. (2007). Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. Journal of cell science *120*, 1245-1255.

Locher, K.P., Lee, A.T., and Rees, D.C. (2002). The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science *296*, 1091-1098.

Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes & development *12*, 1986-1997.

Losada, A., Hirano, M., and Hirano, T. (2002). Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. Genes & development *16*, 3004-3016.

Losada, A., Yokochi, T., and Hirano, T. (2005). Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts. Journal of cell science *118*, 2133-2141.

Losada, A., Yokochi, T., Kobayashi, R., and Hirano, T. (2000). Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. The Journal of cell biology *150*, 405-416.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437-450.

Mc Intyre, J., Muller, E.G., Weitzer, S., Snydsman, B.E., Davis, T.N., and Uhlmann, F. (2007). In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae. The EMBO journal *26*, 3783-3793.

McAleenan, A., Cordon-Preciado, V., Clemente-Blanco, A., Liu, I.C., Sen, N., Leonard, J., Jarmuz, A., and Aragon, L. (2012). SUMOylation of the alpha-Kleisin Subunit of Cohesin Is Required for DNA Damage-Induced Cohesion. Current biology : CB.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of applied crystallography *40*, 658-674.

McGuinness, B.E., Hirota, T., Kudo, N.R., Peters, J.M., and Nasmyth, K. (2005). Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. PLoS biology *3*, e86.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35-45.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D.*, et al.* (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89-102.

Miyazaki, W.Y., and Orr-Weaver, T.L. (1992). Sister-chromatid misbehavior in Drosophila ord mutants. Genetics *132*, 1047-1061.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2006). PCNA controls establishment of sister chromatid cohesion during S phase. Molecular cell *23*, 723-732.

Murray, A.W., and Szostak, J.W. (1985). Chromosome segregation in mitosis and meiosis. Annual review of cell biology *1*, 289-315.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta crystallographica Section D, Biological crystallography *53*, 240-255.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K.*, et al.* (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell *143*, 737-749.

Oikawa, K., Ohbayashi, T., Kiyono, T., Nishi, H., Isaka, K., Umezawa, A., Kuroda, M., and Mukai, K. (2004). Expression of a novel human gene, human wings apart-like (hWAPL), is associated with cervical carcinogenesis and tumor progression. Cancer research *64*, 3545-3549.

Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F., and Nasmyth, K. (2000). Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Current biology : CB *10*, 1557-1564.

Petronczki, M., Matos, J., Mori, S., Gregan, J., Bogdanova, A., Schwickart, M., Mechtler, K., Shirahige, K., Zachariae, W., and Nasmyth, K. (2006). Monopolar attachment of sister kinetochores at meiosis I requires casein kinase 1. Cell *126*, 1049-1064.

Rankin, S., Ayad, N.G., and Kirschner, M.W. (2005). Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. Molecular cell *18*, 185-200.

Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001). Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature *410*, 955-959.

Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science *321*, 563-566.

Rowland, B.D., Roig, M.B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouet, F., Underwood, P., Metson, J., Imre, R.*, et al.* (2009). Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. Molecular cell *33*, 763-774.

Salic, A., Waters, J.C., and Mitchison, T.J. (2004). Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. Cell *118*, 567-578.

Schmitz, J., Watrin, E., Lenart, P., Mechtler, K., and Peters, J.M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Current biology : CB *17*, 630-636.

Sheldrick, G.M. (2008). A short history of SHELX. Acta crystallographica Section A, Foundations of crystallography *64*, 112-122.

Shintomi, K., and Hirano, T. (2009). Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. Genes & development *23*, 2224-2236.

Sonoda, E., Matsusaka, T., Morrison, C., Vagnarelli, P., Hoshi, O., Ushiki, T., Nojima, K., Fukagawa, T., Waizenegger, I.C., Peters, J.M.*, et al.* (2001). Scc1/Rad21/Mcd1 is required for sister chromatid cohesion and kinetochore function in vertebrate cells. Developmental cell *1*, 759-770.

Spell, R.M., and Holm, C. (1994). Nature and distribution of chromosomal intertwinings in Saccharomyces cerevisiae. Molecular and cellular biology *14*, 1465-1476.

Strom, L., Karlsson, C., Lindroos, H.B., Wedahl, S., Katou, Y., Shirahige, K., and Sjogren, C. (2007). Postreplicative formation of cohesion is required for repair and induced by a single DNA break. Science *317*, 242-245.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., and Peters, J.M. (2000). Characterization of vertebrate cohesin complexes and their regulation in prophase. The Journal of cell biology *151*, 749-762.

Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., and Peters, J.M. (2002). The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. Molecular cell *9*, 515-525.

Sundin, O., and Varshavsky, A. (1980). Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. Cell *21*, 103-114.

Takahashi, T.S., Basu, A., Bermudez, V., Hurwitz, J., and Walter, J.C. (2008). Cdc7-Drf1 kinase links chromosome cohesion to the initiation of DNA replication in Xenopus egg extracts. Genes & development *22*, 1894-1905.

Tang, Z., Shu, H., Qi, W., Mahmood, N.A., Mumby, M.C., and Yu, H. (2006). PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. Developmental cell *10*, 575-585.

Tang, Z., Sun, Y., Harley, S.E., Zou, H., and Yu, H. (2004). Human Bub1 protects centromeric sister-chromatid cohesion through Shugoshin during mitosis. Proceedings of the National Academy of Sciences of the United States of America *101*, 18012-18017.

Toth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., and Nasmyth, K. (1999). Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. Genes & development *13*, 320-333.

Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature *400*, 37-42.

Uhlmann, F., and Nasmyth, K. (1998). Cohesion between sister chromatids must be established during DNA replication. Current biology : CB *8*, 1095-1101.

Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V., and Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell *103*, 375-386.

Unal, E., Arbel-Eden, A., Sattler, U., Shroff, R., Lichten, M., Haber, J.E., and Koshland, D. (2004). DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. Molecular cell *16*, 991-1002.

Unal, E., Heidinger-Pauli, J.M., Kim, W., Guacci, V., Onn, I., Gygi, S.P., and Koshland, D.E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. Science *321*, 566-569.

Unal, E., Heidinger-Pauli, J.M., and Koshland, D. (2007). DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). Science *317*, 245-248.

van Heemst, D., James, F., Poggeler, S., Berteaux-Lecellier, V., and Zickler, D. (1999). Spo76p is a conserved chromosome morphogenesis protein that links the mitotic and meiotic programs. Cell *98*, 261-271.

Verni, F., Gandhi, R., Goldberg, M.L., and Gatti, M. (2000). Genetic and molecular analysis of wings apart-like (wapl), a gene controlling heterochromatin organization in Drosophila melanogaster. Genetics *154*, 1693-1710.

Waizenegger, I., Gimenez-Abian, J.F., Wernic, D., and Peters, J.M. (2002). Regulation of human separase by securin binding and autocleavage. Current biology : CB *12*, 1368-1378.

Waizenegger, I.C., Hauf, S., Meinke, A., and Peters, J.M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell *103*, 399-410.

Weitzer, S., Lehane, C., and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. Current biology : CB *13*, 1930-1940.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T.*, et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

Williams, G.J., Williams, R.S., Williams, J.S., Moncalian, G., Arvai, A.S., Limbo, O., Guenther, G., SilDas, S., Hammel, M., Russell, P.*, et al.* (2011). ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. Nature structural & molecular biology *18*, 423-431.

Xu, Z., Cetin, B., Anger, M., Cho, U.S., Helmhart, W., Nasmyth, K., and Xu, W. (2009). Structure and function of the PP2A-shugoshin interaction. Molecular cell *35*, 426-441.

Yamagishi, Y., Sakuno, T., Shimura, M., and Watanabe, Y. (2008). Heterochromatin links to centromeric protection by recruiting shugoshin. Nature *455*, 251-255.

Zhang, N., Panigrahi, A.K., Mao, Q., and Pati, D. (2011). Interaction of Sororin protein with polo-like kinase 1 mediates resolution of chromosomal arm cohesion. The Journal of biological chemistry *286*, 41826-41837.

(1994). The CCP4 suite: programs for protein crystallography. Acta crystallographica Section D, Biological crystallography *50*, 760-763.